

REMARKS

Claims 1-4, 6, 16-17, 27, 29, 34, and 50 are amended herein. Claims 5, 18 and 19 are canceled. New claims 55-56 are added.

Support for the amendment of claim 1 can be found throughout the specification, specifically on page 17, lines 1-10, page 10, lines 18-24 and page 42, lines 15-27. Support for the amendment of claim 3 can be found in the specification at page 33, lines 15-24. Support for the amendment of claim 4 can be found throughout the specification, such as on page 10, lines 18-24 and page 42, lines 15-27. Claim 6 is amended to correct form. Support for the amendment of claim 17 can be found throughout the specification, specifically on page 17, lines 1-10. Claim 18 is amended to correct dependency. Claim 34 is amended to correct a typographical error. Claim 50 is amended to correct form.

Claims 1-3, 16, 17, 27, 29, and 50 are amended to remove subject matter. Applicants reserve the right to pursue the subject matter in a continuation application.

Support for new claim 55 can be found throughout the specification at page 10, lines 7-14, Fig. 14, and page 17, lines 1-10, page 10, lines 18-24 and page 42, lines 15-27. Support for new claims 56-58 can be found throughout the specification, for example on page 24, lines 1-20, pages 25-35.

Applicants believe no new matter is added. Reconsideration of the subject application is respectfully requested.

Objections to the Specification

Applicants thank the Examiner for noting the typographical error in the amendment dated January 11, 2002. The specification should be amended on page 59, and not on page 53. As requested in the Office action, the specification is amended herein to introduce the sequence identifiers in the paragraph on page 59. Upon review of this section of the specification, it was noted that a reference is included in this paragraph to an internet website address. Thus, to expedite prosecution, the specification is also amended to remove a reference to the specific website address for the proteomics server, and to include a statement that this server is "available on the internet."

Restriction Requirement

Applicants thank the Examiner for entering the preliminary amendments, and for the telephone conference of June 7, 2004. Applicants confirm the election of Group I (claims 1-6, 10, 15-20, 24-28, 34, 35, and 45-47), with traverse. Applicants submit that it would not be an undue burden on the Examiner to search the subject matter of Group I with Group V. The reasons for traverse were discussed in the telephone conference of June 7, 2004. However, solely to advance prosecution, the non-elected claims are canceled herein.

Priority

The Office action alleges that claims 1-3, 6, 10, 15-17, 20, 24-28, 34, 35 and 47 are not entitled to the priority date of the PCT/US00/19039, as claim 1 is drawn to a variant including a conservative substitution. Applicants respectfully disagree with this rejection.

The present application is a § 371 U.S. national stage of PCT/US00/19039 filed July 12, 2000, which was published in English under PCT Article 21(2). As required, the specification of the present application is identical to the PCT application (for the Examiner's convenience, a copy of the PCT publication was submitted when the present application was filed, including the coversheet showing the publication number). As discussed in the MPEP 1893.03(b),

“An international application designating the U.S. has two stages (international and national) with the filing date being the same in both stages.It should be borne in mind that the filing date of the international stage application is also the filing date for the national stage application.”

The Notice of Acceptance, dated October 31, 2003, clearly shows that this application entered the national state under 35 USC § 371, and awards the present application the international filing date of July 12, 2000, and the priority date of July 13, 1999.

Applicants note that conservative substitutions are described in the specification of the present application (which is identical to the PCT application) on page 18, lines 14-25. The relationship of conservative substitutions to sequence identity is also disclosed therein.

However, Applicants also note that the claims have been amended to remove reference to variants. Applicants believe that these arguments, and the amendment of the claims should remove any possibility of an allegation that the claims are not entitled to the benefit of the International filing date.

The Office action further alleges that claims 1-3, 6, 10, 15-17, 20, 24-28, 34, 35, and 47 are not entitled to the priority date of the either provisional application as being drawn to polypeptides that bind to an antibody that binds to SEQ ID NO: 14 or as being drawn to polypeptides that are processed and presented in the context of MHC and activates T cells. Applicants respectfully disagree with this assertion.

U.S. Provisional Application No. 60/157,471, filed on October 1, 1999, contains a complete description of immunogenic fragments. This provisional application is entitled "T Cell Receptor Gamma Transcript in Prostate Epithelial Cells," and provides an alternate name for TARP, prostate specific TCR- γ (PS-TCR γ). The specification of this provisional application describes immunogenic fragments in detail. For example, fragments of TARP are described on page 19 of the parent provisional application; lines 1-18 discloses fragments of "at least 5 to at least 15 consecutive amino acids" (line 2) that can "bind to antibodies raised against PS-TCR γ protein" (line 3), or comprise "an epitope that bind an MHC molecule" (line 14). Specific characteristics of these PS-TCR γ polypeptides are also disclosed in the provisional application (such as that they are about 9 or 10 amino acids in length and have a leucine or methionine in the second position, see page 20, lines 6-22). As such, Applicants believe that claims 4 and 18 are entitled to the filing date of the parent provisional application. Methods for identifying antigenic epitopes are also disclosed in the provisional application (for example, see page 21, lines 1-9). As such, Applicants believe that the claims are fully enabled by the provisional application, and should be awarded the filing date of the provisional application.

However, solely to advance prosecution, the claims have been amended to refer to immunogenic fragments including eight to ten consecutive amino acids of SEQ ID NO: 14.

With regard to claims 4 and 18, the Office action alleges that immunogenic fragments that induce a T cell or a B cell response are not entitled to the filing date of either provisional application. Applicants note that U.S. Provisional Application No. 60/157,471, filed on October

1, 1999, contains a complete description of immunogenic fragments. This provisional application is entitled "T Cell Receptor Gamma Transcript in Prostate Epithelial Cells," and provides an alternate name for TARP, prostate specific TCR- γ (PS-TCR γ). The specification of this provisional application describes immunogenic fragments in detail. For example, fragments of TARP are described on page 19 of the parent provisional application; lines 1-18 discloses fragments of "at least 5 to at least 15 consecutive amino acids" (line 2) that can "bind to antibodies raised against PS-TCR γ protein" (line 3), or comprise "an epitope that bind an MHC molecule" (line 14). Specific characteristics of these PS-TCR γ polypeptides are also disclosed in the provisional application (such as that they are about 9 or 10 amino acids in length and have a leucine or methionine in the second position, see page 20, lines 6-22). As such, Applicants believe that claims 4 and 18 are entitled to the filing date of the parent provisional application. Methods for identifying antigenic epitopes are also disclosed in the provisional application (for example, see page 21, lines 1-9). As such, Applicants believe that claims 4 and 18 are fully enabled by the provisional application, and should be awarded the filing date of the provisional application.

With regard to claims 5 and 19, the Office action alleges that these claims are not entitled to the filing date of either provisional application as there is no support for peptides that can be presented in the context of MHC. Applicants respectfully disagree with this assertion.

As discussed above, U.S. Provisional Application No. 60/157,471, filed on October 1, 1999, contains a complete description of immunogenic fragments. For example, page 19 discloses fragments of at least 5 to at least 15 consecutive amino acids of PS-TCR γ (TARP) that are immunogenic that include an epitope that can bind MHC. The provisional application further describes polypeptides of use in binding MHC, of eight to ten amino acids in length, such as that sequences that are about 9 or 10 amino acids in length and have a leucine or methionine in the second position (for example, see page 20, lines 6-22). The specification further contains an assay for identifying antigenic epitopes of use using tumor infiltrating lymphocytes, which can be used to screen epitopes of interest (see page 21 of the provisional application).

Programs for predicting MHC binding motifs were well known to those of skill in the art when the provisional application was filed. For example, TEPITOPE (Sturniolo et al., *Nat. Biotechnol.* 17:555-561, June, 1999, abstract enclosed as Exhibit A) was a matrix based

computer program available at the time the application was filed that was used successfully in locating T cell epitopes in several antigens (see Manici et al., *J. Exp. Med.* 189(5):871-9, March 1999, abstract enclosed as Exhibit B). Additional programs were also available at the time the provisional applications were filed. For example, Brusic et al. (*Bioinformatics* 14(2):121-130, 1998, copy enclosed as Exhibit C) discloses a program (PERUN) that combines a high accuracy of predictions with the ability to integrate new data. Similarly, Atuvia (*Human Immunol.* 58:1-11, 1997, abstract enclosed as Exhibit D) discloses a structure based algorithm that can predict the binding of peptides to MHC molecules and is disclosed to be “a useful tool in the rational design of peptide vaccines...” Thus, it is clear that one of skill in the art, using the PS-TCR γ (TARP) sequence, and the guidance provided by the provisional application, could identify the epitopes of interest at the time the provisional applications were filed.

The Office action further alleges that claim 28, drawn to a method of inducing an immune response using the polypeptide of claim 1 in conjunction with an adjuvant, is not entitled to the filing of the parent PCT application or the provisional applications. Applicants respectfully disagree with this assertion.

As discussed above, the present application is a § 371 U.S. national stage of PCT/US00/19039 filed July 12, 2000. This application is a U.S. national stage application. MPEP 1893.03(b) states that “an international application designating the United States has two stages (international and national), with the filing date being the same in both stages.” Thus, claim 28 cannot be just entitled to the date of entry into the national phase; it must be entitled to the filing date of the international application.

In making this rejection, the Office action states that the adjuvant must be one of the adjuvants and factors listed. Applicants do not agree with the rejection. However, solely to advance prosecution, claim 28 is amended to be in Markush format. This is supported in present specification (which is identical to the PCT specification at page 6, lines 5-9) states the following:

“Additionally, the methods may comprise co-administering to the subject an immune adjuvant selected from non-specific immune adjuvants, subcellular microbial products and fractions, haptens, immunogenic proteins, immunomodulators, interferons, thymic hormones and colony stimulating factors.”

Applicant submits that this amendment removes the objection.

As the Office action includes a detailed review of the support in the specification and the parent applications, it is the Applicants understanding that any claim not specifically addressed in the first Office action is entitled to the benefit of the parent applications.

Oath/Declaration

The declaration was objected to as Drs. Vasmatzis and Wolfgang altered there addresses, but did not initial these changes. Submitted herewith is a new copy of declaration signed by Drs. Vasmatzis and Wolfgang. Applicants believe that the submission of this declaration removes the objection.

Specification

The specification is objected to for not referring to the correct sequences in the legend of Figure 14. This legend is amended herein to refer to the proper sequence identifiers (SEQ ID NOs: 16-18). In the telephone conference with Examiner Rawlings it was discussed that this amendment would overcome the objection. It was further discussed that an additional sequence listing was not required to overcome this objection, as the sequences were included in the original paper copy of sequences listing (and the corresponding CRF).

The specification is objected to for including internet addresses. The specification is amended herein to remove the executable code.

The specification objected to for including trademarks without proper identification. The specification is amended herein to properly refer to trademarks, such as GENBANK® and FASTTRACK.™

In addition, Applicants have corrected typographical errors found in the specification.

No new matter is added. Applicants believe that the amendments remove the objections.

Objection of Claim 34

Claim 34 was objected to for including a typographical error. Claim 34 is amended herein to correct the typographical error. This amendment should remove the objection.

Rejections Under 35 U.S.C. § 112, first paragraph

Claims 26-28 and 34-35 were rejected under 35 U.S.C. § 112, first paragraph, as allegedly the specification does not provide support for “a female at risk for developing breast cancer.” The Office action alleges this is new matter.

Applicants respectfully disagree with this rejection. The specification clearly describes the administration of the peptides to women at risk of developing breast cancer. For example, the specification states (see page 28) that the compositions can be administered to “women prophylactically to provide an immune defense in the event that a TARP-expressing breast cancer develops later.” The specification further states that a “prophylactic” treatment is a “treatment administered to a subject who does not exhibit signs of a disease or exhibits only early signs for the purpose of decreasing the risk of developing pathology” (see page 23, lines 11-14).

Inserting the definition of prophylactic found in the specification into the teachings at page 28, one arrives at the conclusion that TARP polypeptides can be administered to woman who does not exhibit signs of breast cancer or exhibits only early signs of breast cancer for the purpose of decreasing the risk of developing breast cancer, to provide an immune defense in the event that a TARP-expressing breast cancer develops later. Clearly, this provides adequate support for the administration of “females” (woman) “at risk of developing breast cancer” (a woman who does not exhibit signs of breast cancer for the purpose of decreasing her risk). Reconsideration and withdrawal of the rejection is respectfully requested.

Claim 27 is rejected as allegedly there is insufficient support for a composition comprising CD8+ cells pulsed with a variant of SEQ ID NO: 14. The Office action notes that there is support for sensitizing antigen presenting cells with a polypeptide having an epitope of

the amino acid sequence set forth as SEQ ID NO:14. As such, it is the Applicants understanding that the object relates to the term “variant.”

Applicants respectfully disagree with this rejection. As discussed above, Applicants believe there is adequate support for conservative variants of SEQ ID NO: 14 and their use in the specification. However, solely to advance prosecution, claim 27 is amended herein to be limited to epitopes of SEQ ID NO: 14, thereby removing the objection.

Claim 28 is rejected as allegedly there is insufficient support for a composition including more than one adjuvant. As discussed above, Applicants believe there is adequate support for the claimed compositions including adjuvants. However, solely to advance prosecution, claim 28 is amended to include a Markush group, which is clearly supported by the specification. Applicants submit that this amendment renders the objection moot.

Claims 1-6, 10, 15-20, 24-28, 34-35 and 47 are rejected under 35 U.S.C. § 112, first paragraph, as allegedly there is insufficient written description for a genus of variants of SEQ ID NO: 14, the members of which vary by at least a conservative substitution. Applicants respectfully disagree with this rejection, and submit that there is adequate support for variants of SEQ ID NO: 14 in the specification. However, solely to advance prosecution, the claims are amended herein to no longer refer to variants that include at least one conservative substitution. Applicants submit that the amendment of the claims renders the rejection moot.

Claims 1-6, 10, 15-20, 24-28, 34-35 and 47 are rejected under 35 U.S.C. § 112, first paragraph as allegedly there is insufficient written description for a polypeptides with at least 90% sequence identity to TARP. Applicants respectfully disagree with this rejection as applied to the claims as amended which are directed to polypeptides with at least 90% sequence identity to TARP, wherein the polypeptide is expressed by prostate cancer cells, breast cancer cells, or both.

SEQ ID NO: 14 is disclosed in the specification (for example, see Fig. 14). The specification clearly discloses polypeptides sharing at least 90% sequence identity to TARP, SEQ ID NO: 14, such as on page 24, lines 20. Computer programs are disclosed that can readily be used to determine sequence identity (see page 21, lines 25 to page 22, line 6).

TARP is only 58 amino acids in length (see Fig. 14). Applicants note that for an amino acid sequence to have at least 90% sequence identity, it must be identical over at least 53 of 58 amino acids of SEQ ID NO: 14. In other words, at most five out of the 58 amino acids of SEQ ID NO: 14 can be substituted in a polypeptide that is at least 90% identical to TARP.

Applicants submit that the specification readily provides sufficient information for one of skill in the art to substitute at most 5 amino acids in a 58 amino acid sequence (SEQ ID NO: 14).

Moreover, the specification at page 58, lines 11-27 and Fig. 14B provides the location of the functional domains of TARP (a leucine zipper region (amino acids 46-49 and amino acids 55-58) and a cAMP and a GMP phosphorylation site (amino acids 19-21 and 20-22), respectively). Fig. 14B shows a comparison of amino acids 42-57 of TARP with DTUP1 and YTUP1, with conserved domains shown in boxed regions. Non-conserved domains (which could be replaced) are also indicated.

In addition, the specification teaches that TARP is expressed in prostate cancer cells and breast cancer cells and is involved in oncogenic transformation, for example, see page 52, lines 7-13, page 55, lines 19-23, and page 60, lines 8-16.

The Office action asserts that as Skolnik et al. (*TIBTECH* 18:34-39, 2000) discloses that assigning functional activities based on sequence is inaccurate because of the multifunctional nature of proteins, the written description cannot provide adequate guidance for the production of polypeptides with 90% sequence identity that possess a given function. Applicants disagree with this assertion. Skolnik et al. discusses that the term "function" has many meanings (see page 34, column 1 in the section entitled "What is protein function?"). Skolnik et al. discloses that a protein's function at the cellular level can be the interaction with other molecules, or its physiological function. However, Skolnik et al. specifically states that "this article...focuses on identifying the biochemical function of a protein given its sequence." Thus, when Skolnik et al. states that "just knowing the structure of the protein is insufficient for prediction of multiple functional site" it is to be interpreted that the structure is insufficient to predict the biochemical function of the protein. In fact, Skolnik et al. teaches that both structure and knowledge regarding the functional sites must be used to identify the molecular biochemical function of a protein of interest. Skolnik et al. suggests that knowledge of both the sequence (such as the sequence of TARP) and the structural domains (such as the domains shown in Fig. 14) will provide a more accurate understanding of the biological function of proteins related to a protein

of interest (such as TARP). As both sequence and structural information are provided in the specification, Applicants fail to understand how Skolnik et al. suggests that the description does not provide evidence that the Applicants were in possession of the claimed proteins.

In addition, Skolnik et al. states that the interaction with other molecules and the physiological function are very different from the biochemical function (see page 34, second column). Thus, the disclosure of Skolnik et al. cannot be construed to suggest that both structural and sequence information are required to understand physiological function (such as the production of an immune response) or phenotypic function (such as expression by a cell type). Thus, one of skill in the art would NOT conclude based on Skolnik et al. that understanding structural domains would be required to predict cellular expression or binding to an antibody. Again, this suggests that Skolnik et al. is irrelevant to the claims.

Bowie et al. discloses that an amino acid sequence determines the shape and function of a protein (see the abstract). Bowie et al. further teaches that “comparison of different sequences with similar messages can reveal key features...” Applicants do not deny that Bowie et al. describes that the relationship of biochemical function to sequence is complex. However, this relates to the active site in a core sequence involved in a biological function (such as a repressor, see Fig. 1), and the folding of this active site. Moreover, Bowie et al. concludes “...it is now possible to use genetic methods to generate lists of allowed substitutions....at least in the short term, it may not be necessary to solve the folding problem for individual protein sequences. Instead, information from sequence can be used.” Thus, Applicants fail to understand how Bowie et al. supports the argument that the disclosure of the specification is inadequate.

In view of the clear guidance provided by the specification, and the amendments to the claims, Applicants submit that there is sufficient written description for polypeptides with at least 90% sequence identity to TARP, as recited in claims 1-6, 10, 15-20, 24-28, 34-35 and 47. Reconsideration and withdrawal of the rejection is respectfully requested.

Claims 1, 3, 6, 10, 15, 17, 20, 24-28, 34, 35, and 47 were rejected under 35 U.S.C. § 112, as allegedly the specification does not provide sufficient written description for antigenic epitopes.

Applicants respectfully disagree with this assertion. The amino acid sequence of TARP is provided as SEQ ID NO: 14. As noted above, this amino acid sequence is 58 amino acids in

length. Functional domains are disclosed in Figure 14. Epitopes are clearly described in the specification. For example, the specification discloses that epitopes of at least 10 consecutive amino acids, and epitopes that are 8-10 amino acids in length and have anchoring residues (see for example, page 17, lines 1-10, page 24, lines 3-9, pages 28-30, page 33, lines 16-24). Specific configurations of use are disclosed, such as wherein the TARP polypeptides is 9 or 10 amino acids in length and has a leucine or methionine in the second position and valine or leucine in the last positions (for example, see page 28, lines 31-33). In addition, biological methods of testing whether an epitope is immunogenic is also provided (for example, see page 17, lines 3-12 and page 30).

Moreover, computer based programs for predicting MHC binding motifs (immunogenic epitopes) were well known to those of skill in the art at the time the provisional application was filed. For example, TEPITOPE (Sturniolo et al., *Nat. Biotechnol.* 17:555-561, June, 1999, abstract enclosed as Exhibit A) was a matrix based computer program available at the time the application was filed that was used successfully in locating T cell epitopes in several antigens (see Manici et al., *J. Exp. Med.* 189(5):871-9, March 1999, abstract enclosed as Exhibit B). Additional programs were also available at the time the provisional applications were filed. For example, Brusic et al. (*Bioinformatics* 14(2):121-130, 1998, copy enclosed as Exhibit C) discloses a program (PERUN) that combines a high accuracy of predictions with the ability to integrate new data. Thus, given the knowledge of one of skill in the art, and the clear guidance provided by the specification, it is clear that the Applicants were in possession of the claimed polypeptides at the time the application was filed.

Greenspan et al. (*Nature Biotechnology* 7:936-937, 1999) is cited as demonstrating the unpredictability of epitopes. However, Greenspan et al. explores the critical assumptions of alanine scanning mutations, which are used to insert mutations into known epitopes to determine the ligand contact residues (see page 936, columns 1-2). Greenspan et al. concludes that to understand the boundaries of a defined epitope (the ligand contact residues within the epitope) a structural characterization of the molecular interface for binding is required (see page 937, second column). Greenspan et al. further concludes that a number of factors contribute to free energy change involved in the formation of a molecular complex formation (such as the association of antibody with an epitope). At best, the disclosure of Greenspan et al. would suggest that molecular studies would need to be done to clearly identify the ligand contact

residues within each TARP epitope described in the present specification. Thus, the disclosure of Greenspan et al. does not negate the disclosure provided by the specification with regard to identifying immunogenic epitopes of TARP.

Thus, Applicants submit that there is adequate descriptive support for the claimed polypeptides. Reconsideration and withdrawal of the rejection is respectfully requested.

Claims 1-6, 10, 15-20, 24-28, 34-35 and 45-47, with regard to immunogenic fragments of TARP and their use, are rejected as allegedly not being enabled by the specification (see pages 14-22 and 25-27 of the Office action). The Office action also alleges that claims 27, 34 and 35, are not enabled by the specification, as “the specification does not describe with any degree of particularity an epitope of the polypeptide of SEQ ID NO: 14 such that the skilled artisan could make peptides comprising an epitope” (see page 27 of the Office action). Applicants respectfully disagree with these assertions as applied to the claims as amended.

The Office action appears to assert the specification is not enabling for the claimed polypeptides, or epitopes thereof, could not be used for any purpose, such as to induce an immune response against a tumor. In support of this position, the Office action cites Broday et al. (*Anticancer Res.* 20:2665-2676, 2000) and Ezzell et al. (*Journal of NIH Res.* 7:46-49, 1995).

Applicants submit that immunogenic epitopes of TARP and their use are fully enabled by the specification. The amino acid sequence of TARP is provided as SEQ ID NO: 14, which is 58 amino acids in length. Functional domains of TARP are disclosed in Figure 14. Immunogenic epitopes of use are also clearly described in the specification. For example, the specification also discloses that epitopes of use are 8-10 amino acids in length and have anchoring residues. Specific configurations of use are disclosed, such as wherein the TARP polypeptides is 9 or 10 amino acids in length and has a leucine or methionine in the second position and valine or leucine in the last positions (see page 28, lines 31-33). In addition, biological methods of testing whether an epitope is immunogenic is also provided (for example, see page 17, lines 3-12 and page 30). Computer based programs for predicting MHC binding motifs (immunogenic epitopes) were well known to those of skill in the art at the time the provisional application was filed.

In support of this assertion, submitted herewith is a declaration of Dr. Pastan, who describes that antigenic epitopes of TARP, such as those clearly described in the present application, have been generated. Thus, these epitopes are fully enabled by the specification.

With regard to the use of these epitopes, the Office action notes that that the specification clearly teaches that TARP could be used as a target for intervention in prostate cancer and TARP expressing breast cancer, as well as a marker for cancer (see page 27 of the Office action). However, the Office action alleges that insufficient guidance is provided as to whether these epitopes would be of use. Although the Office action concedes that some epitopes could be of use to stimulate a CTL response, it alleges that one of skill in the art could not predict which epitopes could be used generate an immune response against breast cancer cells. Applicants respectfully disagree.

As discussed above, the specification clearly teaches that the epitopes must be at least five amino acids in length. Moreover, the specification teaches epitopes of use are 8-10 amino acids in length and have anchoring residues. Specific configurations of use are disclosed, such as wherein the TARP polypeptides is 9 or 10 amino acids in length and has a leucine or methionine in the second position and valine or leucine in the last positions (see page 28, lines 31-33).

The enclosed declaration of Dr. Pastan provides evidence that the methods described in the present specification were used to generate epitopes of TARP. Specifically, two epitopes were produced. These immunogenic epitopes are nine amino acids in length, have a leucine in the second and last position, and bind HLA (see page 28 of the specification, lines 31-33). The epitopes were used to produce an immune response against a TARP-expressing tumor cell (see the specification at page 27-35). Dr. Pastan's declaration provides evidence that the specification is fully enabling for immunogenic fragments of TARP and their use in generating an immune response against tumor cells.

The Office action also questions that the findings that TARP is expressed by the prostate cancer cell line LNCaP but not PC3. The Office action further questions the data presented in Wolfgang et al. (*Cancer Research* 61:8122-8216, 2001), and suggests that this reference provides support for such an assertion (see the Office action, pages 23-25).

Wolfgang et al. describes that TARP is detected in the androgen-sensitive LNCaP prostate cancer cell line but not in the androgen-independent PC3 prostate cancer cell line (see Fig. 1 of Wolfgang et al.). As discussed in Wolfgang et al., this expression pattern indicated that TARP was important in cancer progression. Thus, to investigate the function of TARP, PC3 cells were produced using an Flp-In-System, which created cells stably transfected with TARP (PC3-TARP cells). To ensure that any detected phenotype was not caused by an integration effect, a cell line was generated harboring the vector without any inserts. To ensure that any detected phenotype was not caused by the nonspecific overexpression of a protein, a cell line that expresses was generated CAT. To ensure that any detected phenotype was not caused by the expression the *TARP* sequence in general, a cell line was generated that expresses the *TARP* gene in the antisense (AS) direction.

To analyze the growth rates, cells derived from each PC3 stable cell line were seeded in triplicate for each time point, and the cell numbers were determined 24, 48, 72, 96 and 120 hours after seeding. Cells were seeded at a density such that they would not reach confluence by 120 hours. The growth medium was not changed to prevent the loss of mitotic cells. Total cell numbers at each time point were determined for each cell line, and their respective doubling times were calculated.

A dramatic increase in growth rate was observed for PC3-TARP cells. PC3-TARP cells had an average doubling time of 16.9 ± 1.3 h, whereas the PC3-Vector, PC3-CAT, and PC3-TARP(AS) cells had average doubling times of 22.6 ± 1.5 , 22.5 ± 1.7 , and 21.5 ± 1.4 h, respectively. Hence, expression of TARP in PC3 cells resulted in a markedly increased growth rate by decreasing their doubling time by more than 5 hours. This data is shown in Fig. 2 of Wolfgang et al. This work provides additional documentation that TARP is involved in the transformation of prostate cancer cells.

To understand the molecular mechanisms behind the increased growth rate observed in PC3-TARP cells, it was investigated whether the expression of TARP in PC3 cells alters gene expression. To do this, the RNA expression profiles of PC3-TARP cells were compared to PC3-Vector cells by cDNA microarray analysis. The data is shown in Fig. 3 of the manuscript. *CAV1*, *CAV2*, *AREG*, and *GRO1* were up-regulated in PC3-TARP cells, whereas *IL-1 β* was down-regulated. Two of the genes found to be induced in the PC3-TARP cells, *AREG* and

CAVI, have been implicated in mediating androgen-stimulated cell growth in prostate cancer cells. As AREG and *CAV1* are androgen-regulated, it was investigated whether TARP was androgen-regulated. To do this, the androgen-sensitive LNCaP cell line was used. LNCaP cells were grown in androgen-depleted media for 48 hours and then treated with either 0.1 nM or 10 nM testosterone at specified time points. The data obtained is presented in Fig. 4. of Wolfgang et al. Testosterone treatment increased *TARP* mRNA levels in the androgen-responsive LNCaP cell line. Thus, the results suggested that *TARP* expression is regulated by androgens. Wolfgang et al. concludes that “these data indicate that TARP may have an important role in prostate cancer...” (see the discussion, first paragraph).

The Office action (see page 23, second paragraph) alleges that because Wolfgang et al. states that because the cell line may have accumulated mutations and epigenetic changes, the results are not significant. Wolfgang et al. does state that PC3 cells have accumulated mutations and epigenetic changes, but suggests that in spite this imperfection in the *in vitro* system, the role of TARP in producing changes in growth rate was still established. The full quote from page 8126 states:

“Presumably, PCS cells have accumulated many mutations and epigenetic changes. *Nevertheless, TARP expression produced changes in growth rate and gene expression in these cells, changes that have been previously described to be predictive of human prostate cancer and to be associated with an increased metastatic potential.*”
[emphasis added]

Thus, Wolfgang et al. concludes that even in this imperfect system, TARP provides an effect associated with the metastatic potential of prostate cancer.

The Office action further alleges that because the Wolfgang et al. publication states that “it is not yet possible to establish the role of TARP in prostate cancer” that the specification simply cannot be enabling for TARP. Applicants respectfully disagree. The full quote is copied below:

“Some lines of evidence suggest that the changes in gene expression observed in PC3-TARP may not be a direct, but instead an indirect effect of TARP expression. For example, TARP is expressed in normal prostate and LNCaP cells (4) . However, *CAV1* is expressed at very low to undetectable levels in normal prostate and LNCaP cells (Ref. 9 and data not shown), and AREG is not expressed in normal prostate (20) . If the induction

of *CAVI* and *AREG* in PC3-TARP cells were a direct effect of TARP expression, one would expect to see *CAVI* and *AREG* expression in normal prostate and *CAVI* expression in LNCaP cells. It is not yet known whether TARP expression results in *CAVI* or *AREG* induction in other prostate cancer cell lines. It is possible that the induction of *CAVI* and *AREG* by TARP may be specific to PC3 cells. Clearly, the molecular mechanisms behind the alteration of gene expression observed in PC3-TARP need additional study.

On the basis of the current results, it is not yet possible to establish the role of TARP in prostate cancer cell growth or normal cell growth. *However, the results presented in this paper propose a pathway that links TARP expression to the modulation of genes involved in generating a malignant phenotype in prostate cancer cells. The question that remains is, what are the downstream components...*” [emphasis added]

When taken in context, it can be seen that Wolfgang is discussing the molecular pathways that include TARP; the downstream components in the biochemical pathways have not yet been established. The present specification establishes the role of TARP in prostate cancer, and Wolfgang et al. describes the role of TARP in promoting cell division, albeit without detailed knowledge of the biochemical pathways within the cell that cause the increase in cell division. Applicants submit that the teachings of Wolfgang et al. simply strengthen the disclosure of the present specification for the role of TARP in tumors. It is because of the clear role of TARP, established by the present application, that the detailed molecular mechanisms should be investigated. In no manner does Wolfgang et al. support any assertion TARP is not involved in cancer, and thus cannot be construed to suggest that the specification is not enabling for the claims.

Claims 1, 4-6, 10, 15, 17, 20, 24-28, 34-35 as directed to a polypeptide with an amino acid sequence at least 90% identical to TARP are rejected as not being enabled by the specification (see the Office action at page 26).

TARP is only 58 amino acids in length (see Fig. 14). Computer based algorithms for the determination of sequence identity are described in the specification, for example at pages 20-21. This information provides sufficient information for one of skill in the art to readily produce the claimed polypeptides.

Applicants note that for an amino acid sequence to have at least 90% sequence identity, it must be identical over at least 53 of 58 amino acids of SEQ ID NO: 14. In other words, at most five out of the 58 amino acids of SEQ ID NO: 14 can be substituted in a polypeptide that is at

least 90% identical to TARP. Applicants submit that the specification readily provides sufficient information for one of skill in the art to substitute at most 5 amino acids in a 58 amino acid sequence (SEQ ID NO: 14).

Moreover, Fig. 14B provides the location of the functional domains of TARP (a leucine zipper region (amino acids 46-49 and amino acids 55-58) and a cAMP and a GMP phosphorylation site (amino acids 19-21 and 20-22, respectively). Fig. 14B shows a comparison of amino acids 42-57 of TARP with DTUP1 and YTUP1, with conserved domains shown in boxed regions. Non-conserved domains (which could be replaced) are also indicated.

Given the guidance provided by the specification, Applicants submit that claims directed to peptides that are 90% identical to TARP and their use are fully enabled. Reconsideration and withdrawal of the rejection is respectfully requested.

Rejections under 35 U.S.C. § 112, second paragraph

Claims 1 and 5, and dependent claims thereof, were rejected for the recitation of “the protein encoded by the amino acid sequence as set forth as SEQ ID NO: 14.” Claim 1 has been amended to remove this phrase. Claim 5 is canceled. Applicants believe that the amendment of claim 1, and the cancellation of claim 5, renders the rejection moot.

Claims 18 and 19 were rejected as being indefinite in depending from canceled claim 12. Claim 18 is amended herein to depend from claim 10. Claim 19 is canceled herein. Applicants believe that the amendment of claim 18, and the cancellation of claim 19, renders the rejection moot.

Rejections under 35 U.S.C. § 102

Claims 1-3, 6, 10, 15-17, 20, 24-25, 27-28 and 47 were rejected as allegedly being anticipated by PCT Publication No. WO 01/04309A1. Applicants respectfully disagree with this rejection.

PCT Publication No. WO 01/04309A1 is the publication of PCT Application No. PCT/US00/19039, which is the international phase of the present application. Thus, the text of PCT Publication No. WO 01/4309 A1 is identical, word-for-word, to the present application.

The Office action alleges that this rejection is made as PCT Application No. PCT/US00/19039 fails to provide an enabling disclosure, and thus the pending claims are not entitled to the benefit of the parent application. As discussed above Applicants disagree with this assertion, and submit that the claims in this national phase application are entitled to the filing date of the parent PCT application.

Indeed, the Office action at page 30 describes the content of PCT Application No. PCT/US00/19039 (the Applicant's own international application), and points out page and line in the specification that provide an example of support for each claimed element. For example, SEQ ID NO: 14 is noted in the Office action to be taught at page 24, lines 1 and 2 of the present specification, amino acid sequences 90% identical to TARP are noted in the Office action to be taught at page 24, lines 3-20 of the present specification, variants that are recognized by antibodies or that bind MHC and can activate T cells expressing SEQ ID NO: 14 are noted in the Office action to be taught at page 5, lines 4-12. The administration of TARP, or TARP polypeptides in a pharmaceutical carrier to a subject with prostate or breast cancer, or who has not been diagnosed with breast cancer are noted in the Office action to be taught in the present specification at page 5, lines 32-34. The administration of TARP polypeptides to induce an immune response, such as with an immune adjuvant, or administering CD8+ cells to an epitope of SEQ ID NO: 14 are noted in the Office action to be disclosed in the present specification on pages 41, lines 25-28, page 6, lines 5-8, and page 6, lines 1-4. Vectors including nucleic acids encoding TARP and variants, including nucleic acids operably linked to a promoter, a noted in the Office action to be disclosed in the present specification at page 26, line 8 to page 27, line 6. The delineation of exemplary support in the specification for each of the claims provides an admission that the present application is indeed entitled to the filing date of the parent PCT application, namely October 1, 1999. Thus, PCT Publication No. WO 01/04309 A1 is not prior art.

In addition, Applicants believe that any rejection over the publication of their own application is improper. The MPEP at 2132.01 states that a 35 U.S.C. § 102(a) *prima facie* case is established only if a reference publication is made "by others." In the present case, the

inventors of this national phase application are identical to inventors of the parent international application. Moreover, MPEP 2132.01 states that any *prima facie* case of anticipation can be rebutted by a showing that the references disclosure was derived from the inventors own work. There is no question that publication of an inventor's own patent application is the work of the same inventor. Indeed, in the present case, the Applicants of the parent PCT application and the Applicants of the U.S. national phase are identical. Thus, no *prima facie* case of anticipation can be established.

Reconsideration and withdrawal of this rejection is respectfully requested.

Claims 1-3, 6, 10, 15-17, 20, 24, 28 and 47 were rejected under 35 U.S.C. § 102(e) as allegedly being rejected by published U.S. Patent Application 2003/0108963 A1 (the '863 application), which allegedly has an effective date of July 25, 2001.

As discussed in detail above, Applicants believe that the pending claims are entitled at least to the benefit of the parent international application, namely July 13, 1999. Indeed the Office action itself notes the support in the parent PCT application for each of the claims. As such, the '863 application is not prior art. Reconsideration and withdrawal of the rejection is respectfully requested.

Applicants believe that the arguments presented above overcome this rejection, and that the '863 application is not prior art. If this rejection is not withdrawn for any reason, the Applicants note that the '863 claims the benefit of four provisional applications. Applicants respectfully request that they be provided with copies of each of the provisional application, so that they can determine the effective date of the '863 application.

Request for a Telephone Interview

Applicants thank Examiner Rawlings for the telephone conference of November 4, 2003, wherein the Office action and the proposed amendments were discussed. Examiner Rawlings indicated that he would review this amendment and discuss the response with a supervisor. Applicants respectfully request an additional telephone interview with the Examiner following entry of this amendment; please contact the undersigned at the telephone number listed below to schedule an interview.

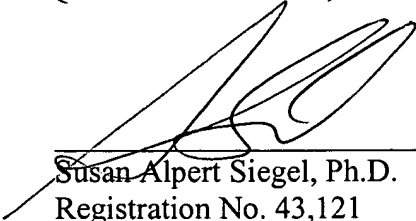
CONCLUSION

If any minor matters remain to be addressed before substantive examination of the application, the examiner is invited to contact the undersigned at the telephone number listed below.

Respectfully submitted,

KLARQUIST SPARKMAN, LLP

By



Susan Alpert Siegel, Ph.D.
Registration No. 43,121

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 226-7391
Facsimile: (503) 228-9446



Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for

Limits Preview/Index History Clipboard Details

 Abstract Text

About Entrez

Text Version

☒ 1: Nat Biotechnol. 1999 Jun;17(6):555-61.

Related Article

Entrez PubMed

Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services

Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Related Resources

Order Documents
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

Comment in:

- Nat Biotechnol. 1999 Jun;17(6):533-4.



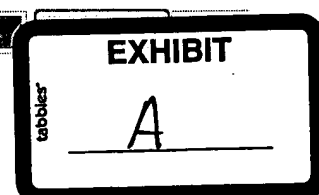
Generation of tissue-specific and promiscuous HLA ligand data using DNA microarrays and virtual HLA class II matrices.

Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxen M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J.

Roche Milano Ricerche, Milan, Italy.

Most pockets in the human leukocyte antigen-group DR (HLA-DR) groove are shaped by clusters of polymorphic residues and, thus, have distinct chemical size characteristics in different HLA-DR alleles. Each HLA-DR pocket can be characterized by "pocket profiles," a quantitative representation of the interaction of all natural amino acid residues with a given pocket. In this report we demonstrate that pocket profiles are nearly independent of the remaining HLA cleft. A small database of profiles was sufficient to generate a large number of HLA-DR matrices, representing the majority of human HLA-DR peptide-binding specificity. These virtual matrices were incorporated in software (TEPITOPE) capable of predicting promiscuous HLA class II ligands. This software, in combination with DNA microarray technology, has provided a new tool for the generation of comprehensive databases of candidate promiscuous T-cell epitopes from human disease tissues. First, DNA microarrays are used to reveal genes that are specifically expressed or upregulated in disease tissues. Second, the predictive software enables the scanning of these genes for promiscuous HLA-DR binding sites. In an example, we demonstrate that starting from nearly 20,000 genes, a database of candidate colon cancer-specific and promiscuous T-cell epitopes can be fully populated within a matter of days. Our approach has implications for the development of epitope-based vaccines.

PMID: 10385319 [PubMed - indexed for MEDLINE]

 Abstract [Write to the Help Desk](#)

Melanoma Cells Present a MAGE-3 Epitope to CD4⁺ Cytotoxic T Cells in Association with Histocompatibility Leukocyte Antigen DR11

By Simona Manici,^{**} Tiziana Sturniolo,[†] Maria Adele Imro,^{**}
Juergen Hammer,^{**} Francesco Sinigaglia,[†] Christoph Noppen,^{†‡}
Giulio Spagnoli,^{†‡} Benedetta Mazzi,[§] Matteo Bellone,^{**}
Paolo Dellabona,^{‡||} and Maria Pia Protti^{**†}

*From the *Laboratory of Tumor Immunology, the †Cancer Immunotherapy and Gene Therapy Program, and the §Laboratory of Molecular Tissue Typing and ||Immunochimistry Unit, Department of Biology and Technology (DIBIT), Scientific Institute H. San Raffaele, 20132 Milan, Italy; †Roche Milano Ricerche, 20132 Milan, Italy; **Roche Discovery Technologies, Hoffman-La Roche, Inc., Nutley, New Jersey 07110; and ‡Departement Chirurgie, Kantonsspital, CH-4031 Basel, Switzerland*

Summary

In this study we used TEPITOPE, a new epitope prediction software, to identify sequence segments on the MAGE-3 protein with promiscuous binding to histocompatibility leukocyte antigen (HLA)-DR molecules. Synthetic peptides corresponding to the identified sequences were synthesized and used to propagate CD4⁺ T cells from the blood of a healthy donor. CD4⁺ T cells strongly recognized MAGE-3₂₈₁₋₂₉₅ and, to a lesser extent, MAGE-3₁₄₁₋₁₅₅ and MAGE-3₁₄₆₋₁₆₀. Moreover, CD4⁺ T cells proliferated in the presence of recombinant MAGE-3 after processing and presentation by autologous antigen presenting cells, demonstrating that the MAGE-3 epitopes recognized are naturally processed. CD4⁺ T cells, mostly of the T helper 1 type, showed specific lytic activity against HLA-DR11/MAGE-3-positive melanoma cells. Cold target inhibition experiments demonstrated indeed that the CD4⁺ T cells recognized MAGE-3₂₈₁₋₂₉₅ in association with HLA-DR11 on melanoma cells. This is the first evidence that a tumor-specific shared antigen forms CD4⁺ T cell epitopes. Furthermore, we validated the use of algorithms for the prediction of promiscuous CD4⁺ T cell epitopes, thus opening the possibility of wide application to other tumor-associated antigens. These results have direct implications for cancer immunotherapy in the design of peptide-based vaccines with tumor-specific CD4⁺ T cell epitopes.

Key words: MAGE-3 • CD4⁺ epitopes • melanoma • tumor vaccines • adoptive immunotherapy

The importance of CD4⁺ T lymphocytes in antitumor immunity has been clearly demonstrated in animal models. CD4⁺ T cells exert helper activity for the induction and maintenance of antitumor CD8⁺ T cells (1-7), but they may also have an effector function either by indirect mechanism against MHC class II-negative tumors, via macrophages activation (for a review, see reference 1), or by direct mechanism against MHC class II-positive tumors (6, 7).

Recently, the requirement of cognate CD4⁺ T cell help for optimal induction of antitumor CD8⁺ CTLs was demonstrated (8). Vaccination with a specific viral T helper epitope, but not with an unrelated T helper epitope, resulted in protective immunity against MHC class II-negative, virus-induced tumor cells. Moreover, simultaneous

vaccination with the tumor-specific T helper and CTL epitopes resulted in strong synergistic protection.

In humans, evidence for a role of CD4⁺ T cells in anti-tumor immunity comes from the study of tumor-infiltrating lymphocytes, which revealed the presence of both CD8⁺ and CD4⁺ T cells at the tumor site (9, 10), and from detection in the sera of neoplastic patients of antibodies directed against tumor antigens (for a review, see reference 11). However, in recent years research on T cell immunity against human tumors has focused mainly on identification of CD8⁺ HLA class I-restricted CTL responses. To date tyrosinase, a tissue-specific antigen expressed in normal and neoplastic cells of melanocytic lineage, is the only melanoma-associated antigen demonstrated as a specific target

for CD4⁺ melanoma-reactive T cells (12, 13) and for which CD4⁺ T cell epitopes have been identified (14).

Characterization of the CD4⁺ T cell epitope repertoire on other tumor-associated antigens, especially those that are tumor-specific and shared among tumors of several histotypes (for a review, see reference 15), would contribute decisively to improve the efficacy of peptide-based immunization protocols in neoplastic patients.

MAGE-3 is a tumor-specific antigen encoded by a gene expressed in a high proportion of melanomas and in several other tumor histotypes (head and neck squamous cell carcinomas, bladder carcinomas, lung carcinomas and sarcomas) and not in normal tissues, with the exception of testis and placenta (for a review, see reference 15). CD8⁺ CTLs from melanoma patients recognize HLA class I-restricted MAGE-3 epitopes (15), and clinical trials with synthetic peptides corresponding to HLA-A1 and/or -A2 MAGE-3 binding sequences are ongoing in patients affected by melanoma and other neoplastic diseases (15). Therefore, MAGE-3 is an excellent candidate protein to study the antitumor CD4⁺ T cell response. This protein has an intracytoplasmic localization (16), making its presentation on MHC class II molecules unlikely or difficult. However, it has been clearly shown that the MHC class II pathway can present endogenous cellular peptides (17–19), and peptides eluted from purified HLA-DR molecules of the melanoma cell line FM3 contained peptides derived from processing of cytoplasmic proteins (20).

In this study, we used a new T cell epitope prediction software (TEPITOPE; reference 21, and our manuscript in preparation) to identify MAGE-3 sequences with promiscuous HLA-DR binding characteristics. Synthetic peptides corresponding to five identified sequences were used to propagate CD4⁺ T cells from the blood of a healthy donor. We show that CD4⁺ T cells are MAGE-3 specific and recognize naturally processed sequence segment(s). Moreover, CD4⁺ T cells are cytolytic and recognize MAGE-3_{281–295} in association with HLA-DR11 on melanoma cells.

Materials and Methods

T Cell Epitope Prediction. TEPITOPE, a new T cell epitope prediction software, is a Windows™ application that enables the identification of (a) class II ligands binding in a promiscuous or allele-specific mode, and (b) the effects of polymorphic residues on class II ligand specificity (21, and our manuscript in preparation). 25 quantitative matrix-based HLA-DR motifs, covering the majority of class II ligand specificity, are incorporated in TEPITOPE (22, and our manuscript in preparation) and provide the basis for various algorithms included in the software package. Starting from any protein sequence, the algorithm permits the prediction and parallel display of ligands for each of the 25 HLA-DR alleles. To predict MAGE-3 CD4⁺ T cell epitopes, we loaded the protein sequence into the software looking for promiscuous peptide regions. We set the TEPITOPE prediction threshold at 5% (21) and picked peptide sequences predicted to bind at least 50% of the HLA-DR molecules incorporated in the software.

DR-Peptide Binding Assay. Peptide interactions with detergent-solubilized DR molecules were measured using an ELISA-based high-flux competition assay (23). HLA-DR molecules were isolated from the following human lymphoblastoid cell lines (LCL): DR1 (DRB1*0101) from HOM-2, DR3 (DRB1*0301) from WT49, DR4 (DRB1*0401) from PREISS, DR5 (DRB1*1101) from SWEIG, DR7 (DRB1*0701) from EKR, and DR8 (DRB1*0801) from BM9. DR2 (DRB1*1501) was isolated from the L cell transfectant L466.1. The molecules were affinity purified using the mAb 1-1C4 (24) as described (25). Peptide competition assays were conducted to measure the ability of unlabeled peptides to compete with a biotinylated indicator peptide for binding to purified DR molecules. The following biotinylated indicator peptides were used: GFK₇ for DR1 and DR7; GIRA₂YA₄ for DR2; LAYDA₅ for DR3; UD₄ for DR4 (26); TT 830–843 for DR5; and GYRA₆L for DR8. The biotinylated indicator peptide and HLA-DR molecules were incubated with 10-fold dilutions (0.001–100 mM) of the unlabeled competitor peptides (peptides corresponding to the MAGE-3 predicted sequences). To determine relative peptide binding affinity, the promiscuous HA_{307–319} peptide from influenza hemagglutinin (27) was included in each competition assay. The relative binding data of the unlabeled competitor peptides were expressed as inhibitory concentration (IC₅₀), i.e., the concentration of competitor peptide required to inhibit 50% of binding of the biotinylated indicator peptide.

Peptide Synthesis. Synthetic peptides corresponding to MAGE-3_{141–155}, MAGE-3_{146–160}, MAGE-3_{156–170}, MAGE-3_{171–185}, and MAGE-3_{281–295} sequences were manufactured on a 9050 Millipore synthesizer. The purity of the peptides was evaluated by reverse-phase HPLC and electron spray mass spectrometry. Synthetic peptides were lyophilized and then reconstituted in DMSO at 2 mg/ml concentration and diluted in PBS as needed.

Cloning and Expression of rMAGE-3. Full-length MAGE-3 coding sequences were inserted into expression vector pET16b (Novagen), allowing the production of the NH₂ terminus 10-histidine tail as described (16). Production and purification of the recombinant fusion protein on nickel column were monitored by SDS-PAGE and Coomassie blue staining.

Propagation of CD4⁺ T Cells. The five synthetic peptides corresponding to the MAGE-3 sequences most promiscuous for HLA-DR binding (see Table I) were pooled (hereafter MAGE-3 pool) and used to stimulate the PBMCs of a healthy donor whose HLA type, identified by standard serologic typing, is A1, A2/B41, B52/DR11, as described (28). In brief, 20 × 10⁶ PBMCs were cultivated for 7 d in RPMI 1640 (GIBCO BRL) supplemented with 10% heat-inactivated human serum (Technogenetics), 2 mM L-glutamine, 100 U/ml penicillin, 50 µg/ml streptomycin (Biowhittaker) (TCM) containing the MAGE-3 pool (1 µg/ml of each peptide). The reactive lymphoblasts were isolated on a Percoll gradient (28), further expanded in T cell growth factor (Lymphocult; Biotest Diagnostic Inc.), and restimulated at weekly intervals with the same amount of antigen plus irradiated (4,000 rad) autologous PBMCs as APCs.

Flow Cytometry. Cytofluorimetric analyses were performed on a FACStarPlus® (Becton Dickinson). The following mAbs were used: anti-CD4-PE and anti-CD8-FITC (Becton Dickinson), D1.12 (purified from an anti-MHC class II hybridoma supernatant), and 57B (described in reference 16). FITC-rabbit anti-mouse Ig antibody (DAKO) was used as second-step reagent in indirect immunofluorescence stainings. Staining for intracytoplasmic MAGE-3 expression was performed as described (29). Intracytoplasmic staining for cytokine expression was performed

using the anti- $\text{INF-}\gamma$ and anti- IL-4 mAbs, following the manufacturer's instructions (Sigma).

Proliferation Assay. CD4^+ T cells and autologous irradiated PBMCs were diluted in TCM to $2 \times 10^5/\text{ml}$ and $2 \times 10^6/\text{ml}$, respectively, and plated in triplicate in 96 round-bottomed well plates (100 μl of CD4^+ T cells and 100 μl of APCs). The cells were stimulated with different concentrations of MAGE-3 pool (0.05, 0.1, 0.5, 1, and 5 $\mu\text{g}/\text{ml}$), each peptide (10 $\mu\text{g}/\text{ml}$), and different concentrations of rMAGE-3 protein (5, 10, and 20 $\mu\text{g}/\text{ml}$). Triplicate wells with CD4^+ T cells alone and APCs alone were used as controls. Three wells with CD4^+ T cells plus APCs did not receive any stimulus in order to determine the basal growth rate (the blank). In inhibition experiments, different concentrations of mAb L243 or an isotype-matched irrelevant mAb (0.25 and 0.5 mg/ml) were added in triplicate wells of CD4^+ cells plus APCs stimulated with MAGE-3 pool (5 $\mu\text{g}/\text{ml}$) or MAGE-3₂₈₁₋₂₉₅ (10 $\mu\text{g}/\text{ml}$). After 3 d, the cultures were pulsed for 16 h with [^3H]TdR (1 mCi/well, 6.7 Ci/mol; Amersham Pharmacia Biotech). The cells were collected with a Titertek multiple harvester (Skatron, Inc.), and the thymidine incorporated was measured in a liquid scintillation counter. The percentage of inhibition was calculated as follows: $[(\text{cpm without mAb} - \text{cpm with mAb})/(\text{cpm without mAb})] \times 100$.

Cytotoxicity Assay. CD4^+ T cells were tested for specific lytic activity in a standard 4-h ^{51}Cr -release assay as described (30). The following targets were used: melanoma cells (SK-Mel 28, HT144, OI TC described in reference 29, and MD TC established in our laboratory from a cutaneous metastasis), and LCL. The HLA-DR type of target cells, identified by molecular or serologic typing, was SK-Mel 28 (DR*04*13), HT144 (DR*04*07), OI TC (DR*01*11), MD TC (DR*04*11), LCL (DR11). In cold target competition assays, unlabeled target cells (cold targets) were seeded in plates at serial ratios of hot-to-cold target cells. Effector CD4^+ T cells and ^{51}Cr -labeled target cells (hot targets) were then added, and cytotoxicity was assessed as described above. Percentage inhibition was calculated as follows: $[(\% \text{ specific lysis without cold target} - \% \text{ specific lysis with cold target})/(\% \text{ specific lysis without cold target})] \times 100$.

Results and Discussion

10 synthetic peptides corresponding to sequence segments predicted by TEPITOPE to form promiscuous MAGE-3 CD4^+ T cell epitopes were synthesized, and their binding to purified molecules of 7 widely diffuse HLA-DR alleles was verified. Based on the results of the competition binding assays, 5 (i.e., the sequences with the greatest degree of promiscuity) of the 10 predicted sequences were chosen for further experiments (Table I). The five synthetic peptides were pooled (MAGE-3 pool) and used to stimulate the PBMCs of a healthy donor. T cells were 94% CD4^+ after 1 wk of culture (not shown), and could be propagated in long-term culture by weekly restimulation with the MAGE-3 pool in the presence of autologous irradiated PBMCs. Reactivity of CD4^+ T cells was tested in microproliferation assays (Fig. 1): the cells responded vigorously to the MAGE-3 pool (Fig. 1 A), even at low concentrations (100–500 ng/ml). Reactivity to the individual peptides forming the pool was also periodically investigated (Fig. 1 C): the CD4^+ T cells recognized predominantly the peptide corresponding to MAGE-3₂₈₁₋₂₉₅

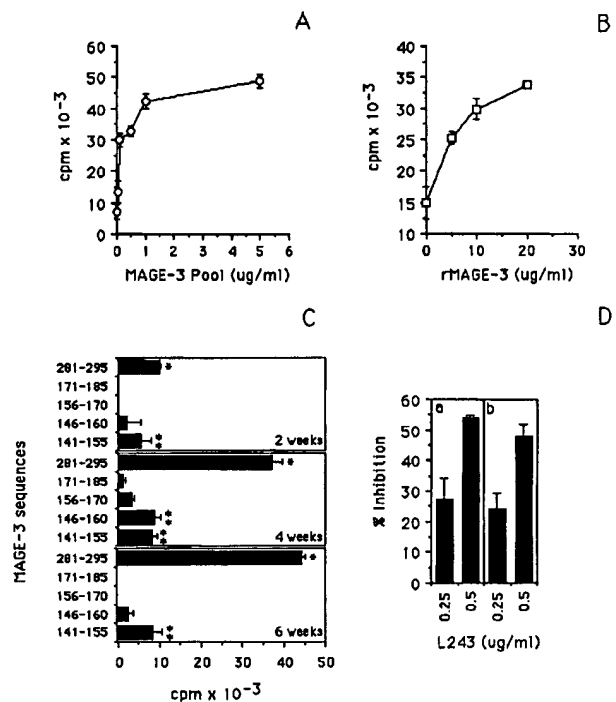


Figure 1. Proliferative activity of CD4^+ T cells stimulated with MAGE-3 pool measured in 2-d microproliferation assays. The data are representative of $n = x$ experiments, and are means of triplicate determinations \pm SD. (A) Responses to MAGE-3 pool (0.01, 0.5, 0.1, 0.5, 1, and 5 $\mu\text{g}/\text{ml}$; $n = 6$). (B) Responses to rMAGE-3 protein (5, 10, and 20 $\mu\text{g}/\text{ml}$; $n = 3$). (C) Responses to the individual synthetic peptides forming the MAGE-3 pool (10 $\mu\text{g}/\text{ml}$; $n = 7$) at different weeks of propagation. The blank (i.e., the basal level of proliferation of CD4^+ T cells in the presence of APCs only) was subtracted and was as follows: 2 wk, $30,866 \pm 1,115$; 4 wk, $7,106 \pm 2,201$; and 6 wk, $21,838 \pm 2,767$. Responses significantly higher than the blanks are indicated as * $P < 0.001$ and ** $P < 0.025$ (determined by unpaired, one-tailed Student's t test). (D) Response to MAGE-3 pool (5 $\mu\text{g}/\text{ml}$; $n = 5$) (a) and to peptide corresponding to sequence 281–295 (b), in the presence of different doses of L243 mAb (0.25 and 0.5 $\mu\text{g}/\text{ml}$). The blank was $1,251 \pm 444$; the proliferation of CD4^+ T cells in the presence of MAGE-3 pool was $28,191 \pm 373$; and the proliferation in the presence of sequence 281–295 was $22,504 \pm 141$.

and, although to a much lower but significant extent, the peptides corresponding to the overlapping sequences MAGE-3₁₄₁₋₁₅₅ and MAGE-3₁₄₆₋₁₆₀. All three sequences recognized by the CD4^+ T cells showed a high binding affinity to purified DR11 molecules (see Table I). Reactivity to MAGE-3₂₈₁₋₂₉₅ increased during the propagation of the line (Fig. 1 C). The proliferative activity of CD4^+ T cells in the presence of MAGE-3 pool (Fig. 1 D, a) or MAGE-3₂₈₁₋₂₉₅ (Fig. 1 D, b) was inhibited by addition in culture of different concentrations of L243 mAb (Fig. 1 D), demonstrating that the recognition of MAGE-3 sequences was HLA-DR restricted. We next tested the CD4^+ T cells for cross-reactivity with the native protein (Fig. 1 B). CD4^+ T cells strongly recognized the rMAGE-3 protein after processing and presentation by autologous APCs, demonstrating that the synthetic sequences recognized by the CD4^+ T cells indeed formed naturally processed epitopes.

Table 1. Determination of HLA-DR Binding of MAGE-3 Synthetic Peptides Corresponding to Sequences Predicted to Form Promiscuous Epitopes

Residues	Sequence	HLA-DR alleles						
		*0101	*0301	*0401	*0701	*0801	*1101	*1501
141–155	GNWQYFFPVIFSKAS	25	>100 [†]	7	0.1	3.2	0.6	3
146–160	FFPVIFSKASSSLQL	10	7	2	0.01	1.5	1.8	0.2
156–170	SSLQLVFGIELMEVD	7	90	45	0.03	7	28	0.18
171–185	PIGHLYIFATCLGLS	0.3	2.8	0.9	0.01	1.5	0.9	0.03
281–295	TSYVKVLHHMVKISG	15	26	70	0.02	0.01	0.03	0.5

The binding data are expressed in terms of relative binding capacity (IC_{50} μ M), calculated as concentration of competitor peptide required to inhibit 50% of the binding of an allele-specific biotinylated peptide (indicator peptide).

[†] IC_{50} values >100 μ M are outside the sensitivity limits of the binding assay.

Intracytoplasmic staining for IL-4 and INF- γ expression, performed after CD4⁺ T cell activation with PMA and ionomycin, revealed that 70% of the CD4⁺ T cells produced INF- γ while no cells produced IL-4 (data not shown), suggesting that they belong mostly to the Th1 type.

To characterize the functional activity of the MAGE-3-specific CD4⁺ T cells, we tested their killing potential against melanoma cells expressing the MAGE-3 protein and the HLA-DR molecules (Fig. 2 B). CD4⁺ T cells showed cytolytic activity against OI TC and MD TC, which express the HLA-DR11 restricting allele, whereas they did not kill SK-Mel 28 and HT144, which express unrelated HLA-DR alleles (Fig. 2 A). To verify whether the cytolytic CD4⁺ T cells recognized HLA-DR11-restricted MAGE-3 epitopes on melanoma cells, we first tested their lytic activity against HLA-DR11⁺ LCL unpulsed, or pulsed with the synthetic peptides recognized in microproliferation assays. LCL pulsed with MAGE-3_{281–295} were strongly recognized by the CD4⁺ T cells, whereas

no killing activity against LCL unpulsed or pulsed with MAGE-3_{141–155} and MAGE-3_{146–160} was detectable (Fig. 3 A). Second, we performed cold target inhibition experiments which showed that the lytic activity of CD4⁺ T cells against OI TC was inhibited by the addition of LCL pulsed with MAGE-3_{281–295} (Fig. 3 B), demonstrating that this sequence is indeed presented by HLA-DR11 on the OI TC melanoma cells. These results further demonstrate that MAGE-3_{281–295} is naturally processed and forms a cytotoxic CD4⁺ T cell epitope. Since the polyclonal CD4⁺ T cells proliferated in the presence of the rMAGE-3 protein, and in addition to MAGE-3_{281–295} they also recognized MAGE-3_{141–155} and MAGE-3_{146–160}, we cannot exclude that these last two sequences may also yield natural epitopes, which are recognized by CD4⁺ T cells with functional activity different from killing. Moreover, although CD4⁺ T cells were mostly Th1 and had direct effector function upon tumor recognition, we cannot exclude that in vivo such CD4⁺ T cells could also exert a helper activity in the induction phase of the immune response.

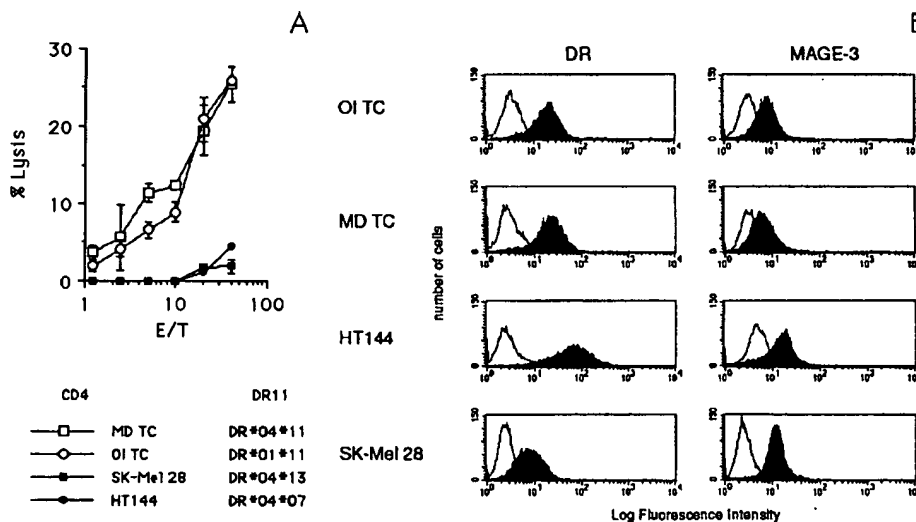


Figure 2. Cytolytic activity of MAGE-3-specific CD4⁺ T cells. The data are representative of $n = x$ experiments, and are means of triplicate determinations \pm SD. (A) Lytic activity against different HLA-DR-matched and -unmatched melanoma cells ($n = 6$). HLA-DR types of CD4⁺ T cells and melanomas are indicated at the bottom along with their symbols. (B) Cytofluorimetric analysis for HLA-DR (surface) and MAGE-3 (intracytoplasmic) expression in melanoma cells used as targets ($n = 4$). Filled histograms, stained sample; open histograms, background staining obtained with FITC-conjugated second-step reagent only.

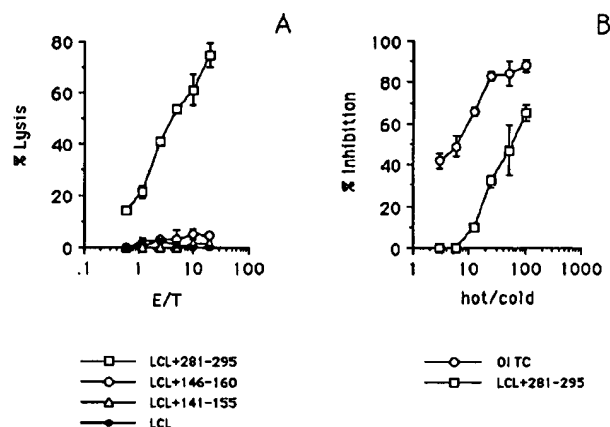


Figure 3. CD4⁺ T cells recognize MAGE-3₂₈₁₋₂₉₅ in association with HLA-DR11 on OI TC cells. The data are representative of $n = 3$ experiments, and are means of triplicate determinations \pm SD. (A) Lytic activity of CD4⁺ CTLs against LCL alone or LCL pulsed with MAGE-3₁₄₁₋₁₅₅, MAGE-3₁₄₆₋₁₆₀, and MAGE-3₂₈₁₋₂₉₅ ($n = 3$). (B) Cold target inhibition experiments ($n = 3$). Cold targets (OI TC [O] and LCL pulsed with MAGE-3₂₈₁₋₂₉₅ [□]) were used to inhibit the lytic activity of MAGE-3-specific CD4⁺ CTLs against hot OI TC (E/T ratio of 40:1). Percentage of specific lysis against OI TC cells in the absence of cold targets was $26 \pm 1.2\%$.

One approach for identifying CD4⁺ T cell epitopes on a candidate protein is the use of overlapping synthetic peptides corresponding to the complete sequence of the protein. The major drawback of this approach is the number of peptide sequences that need to be tested, thus making this approach too expensive and time consuming. In this study, we used the TEPITOPE software package to computationally identify promiscuous HLA-DR binding sites starting from primary protein structures. We demonstrated that TEPITOPE predicted sequence segments capable of binding to multiple HLA-DR alleles. Furthermore, we

showed that one or more of the predicted HLA-DR ligands were indeed naturally processed, thus confirming the validity of this approach. We expect that the application of TEPITOPE to other tumor-associated antigens will speed up identification of the antitumor CD4⁺ T cell epitope repertoire in humans.

Clinical trials based on the use of melanocyte-specific antigens (such as gp100, MART-1/Melan-A, and tyrosinase, for which CD4⁺ T cell epitopes were identified) are in progress in melanoma patients, and although no significant side effects were reported in a recent study that used a gp100 peptide for the treatment of HLA-A2⁺ patients (31), the development of autoimmune responses against normal tissue must be considered when using self-differentiation antigens as vaccines. The demonstration that MAGE-3 (i.e., an antigen not expressed in normal tissues, with the exception of testis and placenta, which are unlikely to be targets of T cells since they do not express MHC molecules), can form CD4⁺ T cell epitopes further supports its use for vaccination protocols in neoplastic patients using a mixture of synthetic peptides corresponding to CD8⁺ and CD4⁺ T cell epitopes.

Previous findings (13, 32, 33) reported a lytic activity of melanoma-specific CD4⁺ T cells. Here we give the molecular definition of an epitope able to stimulate cytolytic CD4⁺ T cells that can be grown in vitro with ease, raising the possibility of using those CD4⁺ T cells in protocols of adoptive transfer in neoplastic patients whose neoplasm expresses the MAGE-3 protein and the MHC class II molecules.

In conclusion, in this study we identified the first CD4⁺ T cell epitope on a tumor-specific antigen, and we verified that the approach used here to predict promiscuous CD4⁺ T cell epitopes yielded natural epitopes. It will be important to evaluate whether the identified CD4⁺ T cell epitopes are indeed promiscuous, making their use for peptide-based vaccines less allele dependent and more widely applicable.

We thank Giuseppe Consogno for excellent technical support and Catia Traversari for critical reading of the manuscript and helpful discussions.

This work was supported by the Fondazione Centro San Raffaele del Monte Tabor, the Italian Association for Cancer Research (AIRC), the Ministry of University and Scientific Research (MURST), and the Swiss National Fund.

Address correspondence to Maria Pia Protti, Laboratory of Tumor Immunology, Scientific Institute H. San Raffaele, Via Olgettina 60, 20132 Milan, Italy. Phone: 39-2-2643-2612; Fax: 39-2-2643-2611; E-mail: m.protti@hsr.it

Received for publication 10 November 1998 and in revised form 15 December 1998.

References

- Greenberg, P.D. 1991. Adoptive T cell therapy of tumors: mechanisms operative in the recognition and elimination of tumor cells. *Adv. Immunol.* 49:281-355.
- Chen, P., and H. Aarathaswamy. 1993. Rejection of K1735 murine melanoma in syngeneic hosts requires expression of MHC class I antigens and either class II antigens or IL-2. *J. Immunol.* 151:244-255.
- Mandelboim, O., E. Vadai, M. Fridkin, A. Katz-Hillel, M. Feldman, G. Berke, and L. Eisenbach. 1995. Regression of established murine carcinoma metastases following vaccination with tumor-associated antigen peptides. *Nat. Med.* 1:1179-1183.

4. Mayordomo, J.I., T. Zorina, W.J. Storkus, L. Zitvogel, C. Celluzzi, L.D. Falo, C.J.M. Melief, S.T. Ildstad, W.M. Kast, A.B. Deleo, and M.T. Lotze. 1995. Bone marrow-derived dendritic cells pulsed with synthetic tumor peptides elicit protective and therapeutic antitumor immunity. *Nat. Med.* 1:1297-1302.
5. Bellone, M., G. Iezzi, A. Martin-Fontecha, L. Rivolta, A.A. Manfredi, M.P. Protti, M. Freschi, P. Dellabona, G. Casorati, and C. Rugarli. 1997. Rejection of a non-immunogenic melanoma by vaccination with natural melanoma peptides on engineered APC. *J. Immunol.* 158:783-789.
6. Ostrand-Roseberg, S., A. Thakur, and V. Clements. 1990. Rejection of mouse sarcoma cells after transfection of MHC class II genes. *J. Immunol.* 144:4068-4071.
7. James, R., S. Edwards, K. Hui, P. Bassett, and F. Grosveld. 1991. The effect of class II gene transfection on the tumorigenicity of the H-2K negative mouse leukemia cell line K36.16. *Immunology.* 72:213-218.
8. Ossendorp, F., E. Mengedé, M. Camps, R. Filius, and C.J.M. Melief. 1998. Specific T helper cell requirement for optimal induction of cytotoxic T lymphocytes against major histocompatibility complex class II negative tumors. *J. Exp. Med.* 187:693-702.
9. Goedegebuure, P.S., and T.J. Eberlein. 1995. The role of CD4⁺ tumor-infiltrating lymphocytes in human solid tumors. *Immunol. Res.* 14:119-131.
10. Maccalli, C., R. Mortarini, G. Parmiani, and A. Anichini. 1994. Multiple sub-set of CD4⁺ and CD8⁺ cytotoxic T-cell clones directed to autologous human melanoma identified by cytokine profiles. *Int. J. Cancer.* 57:56-62.
11. Sahin, U., O. Tureci, and M. Pfreundschuh. 1997. Serological identification of human tumor antigens. *Curr. Opin. Immunol.* 9:709-716.
12. Topalian, S.L., L. Rivoltini, M. Mancini, N.R. Markus, P.F. Robbins, Y. Kawakami, and S.A. Rosenberg. 1994. Human CD4⁺ T cells specifically recognize a shared melanoma-associated antigen encoded by the tyrosinase gene. *Proc. Natl. Acad. Sci. USA.* 91:9461-9465.
13. Yee, C., M.J. Gilbert, S.R. Riddell, V.G. Brichard, A. Fefer, J.A. Thompson, T. Boon, and P.D. Greenberg. 1996. Isolation of tyrosinase-specific CD8⁺ and CD4⁺ T cell clones from the peripheral blood of melanoma patients following in vitro stimulation with recombinant vaccinia virus. *J. Immunol.* 157:4079-4086.
14. Topalian, S.L., M.I. Gonzales, M. Parkhurst, Y.F. Li, S. Southwood, A. Sette, S.A. Rosenberg, and P.F. Robbins. 1996. Melanoma-specific CD4⁺ T cells recognize nonmutated HLA-DR-restricted tyrosinase epitopes. *J. Exp. Med.* 183:1965-1971.
15. Van den Eynde, B.J., and P. van der Bruggen. 1997. T cell defined tumor antigens. *Immunol. Today.* 9:684-693.
16. Kocher, T., E. Schultz-Tjater, F. Gudat, C. Schaefer, G. Casorati, A. Juretic, T. Willmann, F. Harder, M. Heberer, and G. Spagnoli. 1995. Identification and intracellular location of MAGE-3 gene product. *Cancer Res.* 55:2236-2239.
17. Nuchtern, J.G., W.E. Biddison, and R.D. Klausner. 1990. Class II MHC molecules can use the endogenous pathway of antigen presentation. *Nature.* 343:74-76.
18. Chen, B.P., A. Madrigal, and P. Parham. 1990. Cytotoxic T cell recognition of an endogenous class I HLA peptide presented by a class II HLA molecule. *J. Exp. Med.* 172:779-788.
19. Chicz, R.M., R.G. Urban, J.C. Gorga, D.A. Vignali, W.S. Lane, and J.L. Strominger. 1993. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.* 178:27-47.
20. Halder, T., G. Pawelec, A.F. Kirkin, J. Zeuthen, H.E. Meyer, L. Kun, and H. Kalbacher. 1997. Isolation of novel HLA-DR restricted potential tumor-associated antigens from the melanoma cell line FM3. *Cancer Res.* 57:3228-3244.
21. Hammer, J., T. Sturniolo, and F. Sinigaglia. 1997. HLA class II binding specificity and autoimmunity. *Adv. Immunol.* 66:67-100.
22. Hammer, J., E. Bono, F. Gallazzi, C. Belunis, Z.A. Nagy, and F. Sinigaglia. 1994. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J. Exp. Med.* 180:2353-2358.
23. Radrizzani, L., T. Sturniolo, J. Guenot, E. Bono, F. Gallazzi, Z.A. Nagy, F. Sinigaglia, and J. Hammer. 1997. Different modes of peptide interaction enable HLA-DQ and HLA-DR molecules to bind diverse peptide repertoires. *J. Immunol.* 159:703-711.
24. Cammarota, G., A. Scheirle, B. Takacs, D.M. Doran, R. Knorr, W. Bannwarth, J. Guardiola, and F. Sinigaglia. 1992. Identification of a CD4 binding site on the β 2 domain of HLA-DR molecules. *Nature.* 356:799-801.
25. Sinigaglia, F., P. Romagnoli, M. Guttinger, B. Takacs, and J.R.L. Pink. 1992. Selection of T-cell epitopes and vaccine engineering. *Methods Enzymol.* 203:370-386.
26. Hammer, J., F. Gallazzi, E. Bono, R.W. Karr, J. Guenot, P. Valsasini, Z.A. Nagy, and F. Sinigaglia. 1995. Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid arthritis association. *J. Exp. Med.* 181:1847-1855.
27. Roche, P.A., and P. Cresswell. 1990. High-affinity binding of an influenza hemagglutinin-derived peptide to purified HLA-DR. *J. Immunol.* 144:1849-1856.
28. Protti, M.P., A.A. Manfredi, C. Straub, X. Wu, J.F. Howard, Jr., and B.M. Conti-Tronconi. 1990. Use of synthetic peptides to establish anti-human acetylcholine receptor CD4⁺ cell lines from myasthenia gravis patients. *J. Immunol.* 144:1711-1720.
29. Imro, M.A., P. Dellabona, S. Manici, S. Heltai, G. Consogno, M. Bellone, C. Rugarli, and M.P. Protti. 1998. Human melanoma cells transfected with the B7-2 co-stimulatory molecule induce tumor-specific CD8⁺ cytotoxic T lymphocytes in vitro. *Hum. Gene Ther.* 9:1335-1344.
30. Protti, M.P., M.A. Imro, A.A. Manfredi, G. Consogno, S. Heltai, C. Arcelloni, M. Bellone, P. Dellabona, G. Casorati, and C. Rugarli. 1996. Particulate naturally processed peptides prime a cytotoxic response against human melanoma in vitro. *Cancer Res.* 56:1210-1213.
31. Rosenberg, S.A., J.C. Yang, D.J. Schwartzentruber, P. Hwu, F.M. Marincola, S.L. Topalian, N.P. Restifo, M.E. Dudley, S.L. Schwarz, P.J. Spiess, et al. 1998. Immunologic and therapeutic evaluation of a synthetic peptide vaccine for the treatment of patients with metastatic melanoma. *Nat. Med.* 4:321-327.
32. Thomas, W.D., and P. Hersey. 1998. CD4 T cells kill melanoma cells by mechanisms that are independent of FAS (CD95). *Int. J. Cancer.* 75:384-390.
33. Takahashi, T., P.B. Chapman, S.Y. Yang, I. Hara, S. Vijayaradhi, and A.N. Houghton. 1995. Reactivity of autologous CD4⁺ T lymphocytes against human melanoma. Evidence for a shared melanoma antigen presented by HLA-DR15. *J. Immunol.* 154:772-779.

Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network

Vladimir Brusic, George Rudy, Margo Honeyman,
Jürgen Hammer¹ and Leonard Harrison

The Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital,
Parkville, Victoria 3050, Australia and ¹Roche Milano Recherche, Milan, Italy

Received on May 19, 1997; revised on August 23, 1997; accepted on August 25, 1997

Abstract

Motivation: Prediction methods for identifying binding peptides could minimize the number of peptides required to be synthesized and assayed, and thereby facilitate the identification of potential T-cell epitopes. We developed a bioinformatic method for the prediction of peptide binding to MHC class II molecules.

Results: Experimental binding data and expert knowledge of anchor positions and binding motifs were combined with an evolutionary algorithm (EA) and an artificial neural network (ANN): binding data extraction → peptide alignment → ANN training and classification. This method, termed PERUN, was implemented for the prediction of peptides that bind to HLA-DR4(B1*0401). The respective positive predictive values of PERUN predictions of high-, moderate-, low- and zero-affinity binders were assessed as 0.8, 0.7, 0.5 and 0.8 by cross-validation, and 1.0, 0.8, 0.3 and 0.7 by experimental binding. This illustrates the synergy between experimentation and computer modeling, and its application to the identification of potential immunotherapeutic peptides.

Availability: Software and data are available from the authors upon request.

Contact: vladimir@wehi.edu.au

Introduction

Major histocompatibility complex (MHC) molecules play a critical role in initiating and regulating immune responses. MHC molecules bind short peptides and display them on the cell surface for recognition by the T-cell receptor (TCR) of T cells (reviewed in Rammensee *et al.*, 1993; Cresswell, 1994; Engelhard, 1994). Binding of a peptide to an MHC molecule is a prerequisite for recognition by the T cells, but only certain peptides can bind to any given MHC molecule. Determining which peptides bind to a specific MHC molecule is fundamental to understanding the basis of

immunity, and for the development of vaccines and immunotherapeutics for autoimmune disease and cancer.

MHC class II molecules bind peptides that are 10–30 amino acids long (Chicz *et al.*, 1993) with a core region of 13 amino acids containing a primary anchor residue (Jardetzky *et al.*, 1996). Analysis of binding motifs (see Rammensee *et al.*, 1995) suggests that only a core of nine amino acids within a peptide is essential for peptide/MHC binding. Class II molecules contain a single primary anchor, which is necessary for binding, and several secondary anchors that affect binding. Experimental testing of a protein to determine which of its peptide subsequences bind to a specific MHC class II molecule requires binding assays of multiple overlapping peptides spanning the length of the protein. Prediction methods for identifying binding peptides could minimize the number of peptides required to be synthesized and assayed, and thereby facilitate the identification of potential T-cell epitopes.

The prediction of MHC class II-binding peptides is a difficult classification problem. Among the difficulties that must be addressed are: (i) the variable lengths of reported binding peptides; (ii) the undetermined core regions for individual peptides; (iii) the number of amino acids permissible as primary anchors; (iv) the range of experimental methods for assaying of peptide binding; (v) the experimental and reporting errors. Several methods have been used to predict MHC binding peptides, including those based on binding motifs, quantitative matrices and artificial neural networks (ANNs). Binding motifs specify which residues at given positions within the peptide are necessary or favorable for binding to a specific MHC molecule. Sette *et al.* (1989) first described allele-specific motifs for two mouse MHC class II molecules, and motifs for various human and mouse MHC class I and class II molecules have been reported subsequently (see Rammensee *et al.*, 1995). Motifs for MHC class I molecules are relatively well defined. Nijman and co-workers (1993) compared experimental results for binding to HLA-A2.1 with those obtained by motif-based prediction. Of 35 predicted binding peptides, they found that only 15 (43%)



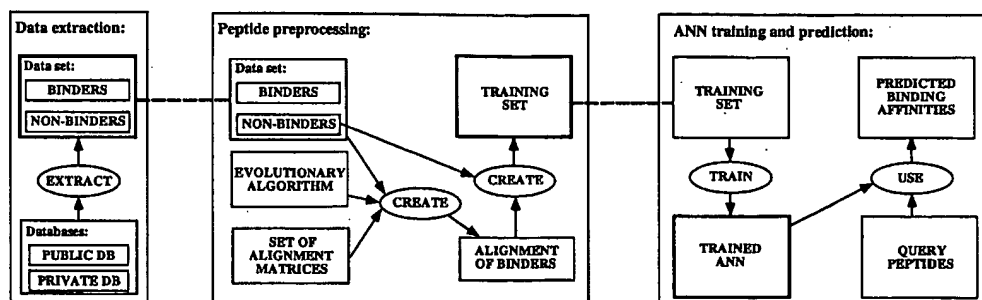


Fig. 1. The overall structure of PERUN. In the data extraction stage, peptide sequences and their binding affinities are collected from a variety of sources. In the pre-processing stage, an evolutionary algorithm generates alignment matrices which are then used to find and align putative nonamer cores of known binders. The ANN training set comprising aligned binding nonamer cores and non-binding nonamers is used in the final stage to train ANNs to predict the binding affinity of query peptides. Dashed lines indicate identity.

actually bound. With the exception of certain molecules (Hammer *et al.*, 1994a; Rothbard *et al.*, 1994; Harrison *et al.*, 1997), specific binding motifs for MHC class II molecules are less well defined (Rammensee *et al.*, 1995).

Quantitative matrices are essentially refined binding motifs. They provide coefficients for each amino acid/position that can be used to calculate scores predictive of binding. The assumptions are that each residue contributes independently of other residues to binding and when located at a given position contributes the same amount to binding even within different sequences. Quantitative matrices have been defined for class I (Parker *et al.*, 1994; Kondo *et al.*, 1995; Schönbach *et al.*, 1995; Brusic *et al.*, 1997) and for class II (Hammer *et al.*, 1994a; Rothbard *et al.*, 1994; Davenport *et al.*, 1995) molecules.

ANNs are connectionist models commonly used for classification (Weiss and Kulikowski, 1990) and pattern recognition (Beale and Jackson, 1990) tasks. ANNs used for the prediction of MHC class I binding peptides have achieved both positive and negative predictive values of nearly 80% (Brusic *et al.*, 1994; Adams and Koziol, 1995). Because of ambiguities resulting from the variable length of reported binders and the uncertain location of their core regions, peptides tested experimentally for binding and used as inputs to train an ANN require pre-processing by alignment relative to their binding anchors. For MHC class I peptides, this is a simple problem because of the presence of well-defined anchor positions and minimal variability in peptide length. MHC class II-binding peptides, however, have more degenerate motifs. Growing evidence (Rammensee *et al.*, 1995) supports the observation by Hammer *et al.* (1993) that MHC class II-binding peptides contain a single primary anchor at the amino terminus, which is a hydrophobic amino acid (Y, F, W, I, V, L or M). The greater variability in length of MHC class II-binding peptides and their less well-characterized

motifs make their alignment a difficult task, particularly as the vast majority contain more than one hydrophobic residue, allowing for multiple possible alignments. Application of a standard multiple alignment method, such as GCG Pileup (<http://www.gcg.com/>), failed to produce a useful alignment. In that alignment, a 9mer core was not preserved, nor would the sequences align relative to the primary anchors.

Each of the described prediction methods has its advantages and drawbacks. Binding motifs encode the most important rules of peptide/MHC interaction, but do not generalize well. Quantitative matrices can predict large subsets of binding peptides reasonably well, but cannot deal with non-linearity within data and may miss distinct subsets of binders. Also, quantitative matrices are not adaptive and self-learning, so that integration of new data usually requires redesigning of the matrix. ANNs can deal with non-linearity and are adaptive and self-learning, but require a large amount of pre-processed data. An ideal prediction method would integrate the strengths of these individual methods while minimizing their disadvantages.

We have therefore developed PERUN, a hybrid method for the prediction of peptides that bind to MHC class II molecules. It utilizes: (i) available experimental data and expert knowledge of binding motifs; (ii) alignment (quantitative) matrices for pre-processing; (iii) an evolutionary algorithm to derive alignment matrices; and (iv) an ANN for classification. The key elements of PERUN are depicted in Figure 1. We have tested the ability of PERUN to predict peptides that bind to HLA-DR4(B1*0401) human MHC class II molecule associated with insulin-dependent diabetes and rheumatoid arthritis, and validated prospectively its predictive accuracy. PERUN combines high accuracy of predictions with the ability to integrate new data and self-improve.

Table 1. Example of an alignment matrix. Each residue at each position in a 9mer is assigned a weighting which is used to calculate a binding score. This particular matrix was derived after 10^6 cycles of reproduction and is characterized by good discrimination between binding and non-binding peptides to HLA-DRB4: classification of known binders was 85% correct and that of non-binders 100%

Amino acid	Amino acid position within the peptide								
	1	2	3	4	5	6	7	8	9
A	-20.0	1.8	0.3	1.1	0.2	0.5	-0.3	1.4	1.1
C	-20.0	2.0	-1.2	1.9	-1.2	-1.8	2.1	-0.1	0.9
D	-20.0	-2.4	-1.9	0.8	-0.8	-0.7	-1.4	-1.8	-1.9
E	-20.0	-1.0	0.7	-2.4	0.2	0.6	0.5	-1.3	-2.2
F	0.0	1.4	-0.6	1.9	0.1	0.4	2.0	-2.3	-1.1
G	-20.0	-1.2	-0.9	-1.2	0.1	1.1	-0.3	0.5	0.3
H	-20.0	-0.6	1.3	0.1	1.2	-0.8	2.0	1.5	-1.0
I	-1.0	-0.7	0.6	1.1	0.5	-1.2	0.6	-0.5	-0.2
K	-20.0	0.4	-2.4	-2.1	0.9	-0.7	0.7	-0.7	-1.8
L	-1.0	-1.8	0.7	0.0	0.1	0.6	0.9	0.2	-2.1
M	-1.0	0.2	-2.1	2.5	0.8	-0.3	2.1	0.5	-1.7
N	-20.0	0.4	-1.7	0.4	-1.0	0.4	1.9	1.8	-1.8
P	-20.0	-0.5	-0.7	0.0	0.3	1.1	0.6	0.2	-1.4
Q	-20.0	-0.2	1.0	1.0	-0.9	-2.0	0.4	0.8	0.1
R	-20.0	2.5	-1.7	0.1	0.5	-0.1	0.4	0.6	-0.8
S	-20.0	0.9	0.2	-1.6	-0.2	0.4	-1.2	1.8	1.2
T	-20.0	1.2	2.1	-1.9	-0.9	2.1	-0.4	2.4	-1.5
V	-1.0	-2.5	1.3	0.7	0.6	0.7	0.9	1.1	1.0
W	0.0	-1.0	0.2	1.1	-2.3	1.5	-1.9	1.8	0.1
X ^a	-20.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Y	0.0	0.2	-0.5	-0.1	-1.1	-1.3	0.0	0.4	1.8

^aX is unknown (any) amino acid.

System and methods

Peptides

Peptide sequences were drawn from MHCPEP, a database of MHC binding peptides (Brusic *et al.*, 1996), from a collection of MHC non-binding peptide data (V.Brusic, unpublished) and from sets of experimental binding data (Hammer *et al.*, 1994b; L.C.Harrison and M.C.Honeyman, unpublished). The initial data set comprised 650 peptides known to bind (338) or not bind (312) to HLA-DR4(B1*0401). The experimental validation set comprised 62 16mer peptides, overlapping by 10 amino acids, spanning the intracytoplasmic domain of human tyrosine phosphatase IA-2, a target of autoimmunity in insulin-dependent diabetes.

Evolutionary algorithm

An evolutionary algorithm (EA) was used to search for predictive peptide alignments. An EA is a search method based on evolutionary principles (Holland, 1975; Goldberg, 1989; Forrest, 1993) in which alternative structures are improved through genetic mechanisms (mutation, cross-over and reproduction) and competition. A population, in this case a set of alignment matrices, is transformed into a new population

(generation) using genetic mechanisms and a selection process for improved fitness. The measure of the predictive power of an alignment matrix was used to define its fitness (see the Algorithm section). The format of matrices was adopted from Hammer *et al.* (1994a) with some modifications (Table 1).

Table 2. Descriptive binding affinities that correspond to ranges of peptide binding affinity determined experimentally. IC_{50} is the concentration of peptide that inhibits binding of a standard peptide by 50%. ANN representation is the output value (associated with each training input to the ANN)

Binding affinity	IC_{50} (μ M)	ANN output
High	≤ 0.9	10
Moderate	1–9	8
Low	10–49	6
None	≥ 50	0

Knowledge of primary anchor positions in reported binding motifs (Rammensee *et al.*, 1995) was used to fix position one (1), corresponding to the primary anchor in each matrix, while the rest of the matrix was subject to the application of the EA. The selection technique was elitist in that each parent (matrix) produced two offspring, an identical copy of itself

and a mutant copy, passing the offspring with the higher fitness value to the next generation. All matrices of the final generation were used to score peptide alignments by assigning a score to each putative binding core within each binding peptide. In each simulation, the alignment scored as highest by the majority of the final generation matrices was selected and passed to the final stage, ANN training. Modifications of a simple evolutionary algorithm (Holland, 1975) used in this work include the use of real number instead of binary representations, omission of cross-over operator (see Discussion) and incorporation of heuristic rules. All programs for the implementation of the EA stage were written in Fortran 77.

Artificial neural networks

An ANN consists of nodes (computational elements) that receive signals via interconnecting arcs. An ANN can be trained to recognize a pattern by strengthening signals (adjusting arc weights) and by adjusting activation thresholds for individual nodes. When trained on a large amount of input data, an ANN can 'extract' and 'remember' generalized patterns present in the data set, and subsequently 'recognize' these patterns in a new, previously 'unseen' input.

The PlaNet package, Version 5.6 (Miyata, 1991), was used to design and train a three-layer fully connected feed-forward ANN (see Zurada, 1992). For all networks, the input layer consisted of 180 nodes, corresponding to the represen-

tation of a nonameric peptide with a single node output layer. Amino acids were represented as binary strings of length 20, of 19 zeros and a unique position set to one for each amino acid. The output value, representing binding affinity, was between 0 and 10. This corresponded to log ranges of binding affinity: 0, no binding; 6, low affinity; 8, moderate affinity; 10, high affinity (Table 2). ANNs with between one and four hidden layer nodes were tested for performance. The learning procedure was error back-propagation (Rumelhart *et al.*, 1986), with a sigmoid activation function (see Zurada, 1992, pp. 41–42). Values for learning rate and momentum were 0.2 and 0.9, respectively. Training was performed with training set randomization in each cycle.

Validation of results

Predictions of binding and non-binding peptides were validated using internal cross-validation as well as by experimental peptide binding. The initial set of 650 peptides was randomly partitioned into training and test sets, the former comprising ~75% of the peptides. Ten such mutually exclusive partitions (Table 3) were used for a 10-fold cross-validation for estimation of the true error rate of the method (described in Weiss and Kulikowski, 1990). The prediction of binding peptides was also validated against the results of direct binding assays on a set of overlapping 16mer peptides from human tyrosine phosphatase IA-2.

Table 3. The composition of cross-validation peptide sets. Peptides listed as unknown have been previously reported as binders, but without binding affinity specified. For ANN training, they were treated as moderate-affinity binders

Sets	Number of peptides grouped by binding affinity					Total
	High	Moderate	Low	Unknown	None	
Set1 train	111	61	67	7	247	493
Set1 test	39	27	22	4	65	157
Set2 train	112	68	70	6	228	484
S2 test	38	20	19	5	84	166
Set3 train	114	70	67	9	226	486
Set3 test	36	18	22	2	86	164
Set4 train	101	67	64	7	225	464
Set4 test	49	21	25	4	87	186
Set5 train	109	74	66	6	230	485
Set5 test	41	14	23	5	82	165
Set6 train	119	69	65	9	216	478
Set6 test	31	19	24	2	96	172
Set7 train	107	66	67	9	244	493
Set7 test	43	22	22	2	68	157
Set8 train	117	70	71	9	229	496
Set8 test	33	18	18	2	83	154
Set9 train	115	72	79	8	235	509
Set9 test	35	16	10	3	77	141
Set10 train	110	68	61	10	217	466
Set10 test	40	20	28	1	95	184

The performance of PERUN was compared to that of a quantitative matrix (Hammer *et al.*, 1994a) and a binding motif (as in Rammensee *et al.*, 1995) weighted as described in Nijman *et al.* (1993). The comparison was performed using Relative Operating Characteristic (ROC) analysis (Swets, 1988). ROC analysis provides a single measure, Aroc, which is a proportion of the area under the ROC—the plot of the true positive proportion versus the false-positive proportion for the various thresholds of the decision criterion. This measure removes biases due to disparate proportions of binding and non-binding peptides, and biases due to arbitrary defined decision thresholds.

Hardware

ANN learning experiments and cross-validation were executed on a Sun Microsystems SPARC 2 4/75 under the SunOS 4.1.3 operating system. Data extraction, pre-processing of ANN input data, pre-processing of cross-validation results and EA searches were performed on a DEC Alpha 3000-400 under the Open VMS V6.1 operating system. Cross-validation and statistical tests were performed on an Apple Macintosh Quadra 800 (System 7.5).

The algorithm

Data extraction

The information extracted from the databases consisted of peptide strings and their experimentally determined binding affinity. The values 0, 6, 8 and 10 corresponding to zero, low-, moderate- and high-affinity binding were used for ANN training. The initial set of 650 peptides was randomly partitioned into training and test sets for a cross-validation, while all 650 peptides were then used for ANN prediction of binding affinities of IA-2 peptides.

Peptide pre-processing

All peptides of known binding affinity were reduced to putative binding nonamer cores or non-binding nonamers. Position one (1) in each nonamer corresponds to the primary anchor. The primary anchor of peptides that bind to HLA-DR(B1*0401) can be any one of the following: I, L, V, M, F, Y or W (see Rammensee *et al.*, 1995). This Set of Allowed Anchor Residues will hereafter be referred to as SAAR. No other amino acid has been observed to serve as a primary anchor for HLA-DR4(B1*0401).

Each non-binder was resolved into as many putative non-binder nonamers as it has positions occupied by SAAR residues (excluding those too close to the C-terminus to yield nonamers). The number of non-binder nonamers derived from peptides in the original set was 578. Each binding peptide yielded a single putative binder, determined using alignment matrices. Reported binders were extended by two positions

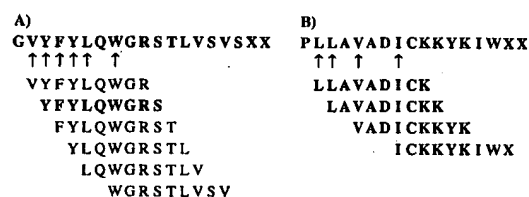


Fig. 2. (A) The DR4 binding peptide GYFYLYLQWGRSTLVSVS (Ig heavy chain 121–137), which has six potential primary anchors, yields a single putative binder YFYLYLQWGRS (shown in bold), while the other five peptides are discarded. (B) Four putative non-binder nonamers are derived from a DR4 non-binding peptide PLLAVADICKKYKIWX (human GAD-65 347–361). The peptide selection process is described in the text.

with 'XX' as necessary to accommodate those that were nonamers and which had an SAAR residue at position two or three rather than at position one. Examples of the resolution of peptides into nonamers are shown in Figure 2.

Putative binders were chosen with alignment matrices (Table 1). An alignment matrix was used to score each nonamer subsequence within the peptide. For example, a known high-affinity binding peptide YRAFATTWQ scores 8.5 ($0 + 2.5 + 0.3 + 1.9 + 0.2 + 2.1 - 0.4 + 1.8 + 0.1$). The score for a peptide is that of its highest scoring nonamer subsequence. The threshold for binding is set to 2.0 (as defined by Hammer *et al.*, 1994a): binders scoring ≥ 2 and non-binders scoring < 2 were considered correctly classified. A population of matrices was initialized and subsequently evolved using EA to determine those matrices that discriminate binders from non-binders. The fitness function was chosen to be $(SE + 3 \times SP)/4$, where $SE = TP/(TP + FN)$ and $SP = TN/(TN + FP)$; SE = sensitivity, SP = specificity, TP = true positives, FN = false negatives, TN = true negatives, FP = false positives. This fitness function favors matrices that correctly classify non-binders and should result in a population of matrices in which individual matrices capture disjoint regions in the solution space. Weights for SAAR at the primary anchor position in each matrix were fixed to 0 (F, Y and W) or -1 (I, L, V and M) as in Hammer *et al.* (1994a). Non-SAAR at the primary anchor position were weighted -20 to disqualify nonamers lacking an anchor. Values for X at all non-anchor positions, representing an unknown amino acid or peptide extension, were set to -1, an arbitrary penalty. All other positions were subject to the application of the EA with the allowed values for these positions between -2.5 and 2.5, adapted from the quantitative matrix of Hammer *et al.* (1994a). The genetic operators used were mutation and reproduction, but not crossover (see Discussion).

Some peptide families (e.g. polyalanine peptides) are over-represented in the data set. To correct for the effect of this bias, peptides were weighted. The weight of a peptide was calcu-

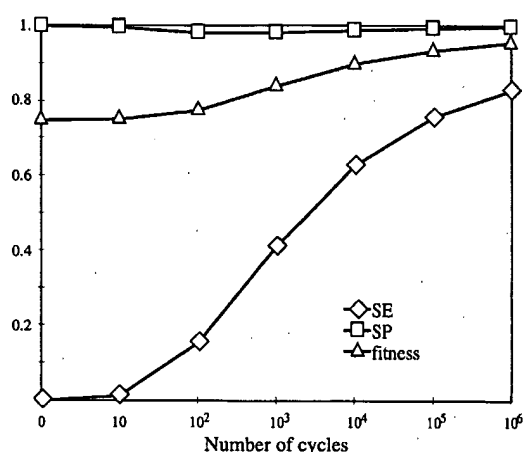


Fig. 3. Sensitivity (SE), specificity (SP) and average fitness function, for a representative EA simulation, of 10 alignment matrices versus the number of search cycles (generations). SE and SP were calculated by using each matrix to classify binding versus non-binding peptides in the primary data set.

lated by taking into account its similarity to other peptides, determined by a simple dot matrix method (Gibbs and McIntyre, 1970) with a specialized scoring matrix (data not shown). These weights ranged from 0.1 for peptides from well-represented families, to 1 for peptides that were dissimilar to others.

Termination criteria for the application of EA were determined in a separate experiment, in which a population of matrices was evolved up to 10^6 generations. Results of a representative experiment are shown in Figure 3. The maximum sensitivity approaches 85%, but this is highly likely to reflect a data overfitting effect. On the basis of these results, 20 000 generations were selected as a termination condition. The final generation of matrices was used to score potential alignments, from which the highest scoring alignment of binding peptides was selected. This alignment, along with putative non-binders, was used for ANN training.

ANN training

ANNs were trained up to 300 cycles using the generalized delta rule (McClelland and Rumelhart, 1986). Architectures containing between one and four hidden layer nodes were tested by internal cross-validation. In addition, a linear ANN which is equivalent to a binding matrix was tested. This range for the number of hidden layer nodes was selected taking into account the complexity of the ANN (total number of arcs and activation thresholds) and the number of training cases (see Discussion).

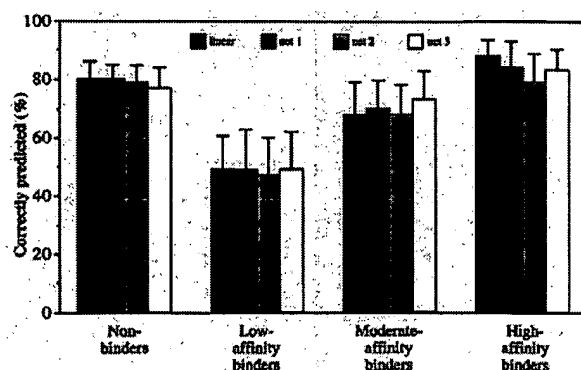


Fig. 4. Summary of cross-validation results grouped by peptide binding affinity and tested ANN topologies. Net1, net2, net3 are topologies with one, two and three hidden layer nodes, respectively. 'Linear' represents a linear network. Mean percentages with one standard error of the mean are shown.

Implementation

Internal cross-validation

The data set was divided into 10 different partitions (groups) as described above. Four different ANN architectures containing between one and four hidden layer nodes were studied. For each group/architecture combination, five training sessions were conducted, resulting in a total of 200 different networks being tested. Training patterns consistently demonstrated the predictable behavior required for accurate analysis (Weiss and Kulikowski, 1990, p. 107): (i) error distances on the training sets decreased with the addition of hidden units; (ii) error distances on replication of the training sessions were reasonably close; (iii) training solutions were as good or better than the alternative methods. Each network was trained up to 300 cycles, a number previously shown to be sufficient for correct learning, but below the overtraining limit for this type of prediction (Brusic *et al.*, 1994). Intermediate results at 50 cycle intervals were recorded and analyzed. Performance was evaluated for both two-class (binders versus non-binders) and four-class (non-binding, low-, moderate- and high-affinity binding) classifications.

The two-class classification (binders versus non-binders) is summarized in Figure 4. For convenience, peptides tested experimentally were separated into groups of non-binders, low-, moderate- and high-affinity binders. The average numbers of peptides in the test groups were: 74 non-binders, 21 low-, 20 moderate- and 38 high-affinity binders. Overall, 80% of non-binders were correctly classified. Approximately 50, 70 and 80% of binders of low, moderate and high affinity, respectively, were correctly classified. The complexity of the network, i.e. the number of hidden layer nodes, did not significantly affect predictive performance.

Table 4. Representative result of four-class classification (average of 50–300 cycles, net3). Values are given as mean percentages with the standard error of the mean in parentheses

Experimental binding affinity	Predicted				Total peptides
	High	Moderate	Low	Non-binder	
High	62 (9)	14 (6)	7 (5)	17 (7)	38
Moderate	41 (13)	22 (12)	10 (8)	27 (10)	20
Low	23 (11)	16 (9)	11 (7)	50 (13)	21
Non-binder	8 (4)	7 (4)	8 (4)	77 (7)	74

A representative matrix for the four-class classification is given in Table 4. Non-binders and high-affinity binders were well classified; low- and moderate-affinity binders were less well classified. This is likely to be due to the arbitrary definition of boundaries between classes, compounded by the smaller number of low- and moderate- than high-affinity binders in the data set.

To establish the minimum number of cycles required for satisfactory ANN training, we observed the classification of high-affinity binders as a function of the number of cycles. Fifty cycles appeared sufficient for training ANNs with 2–4 hidden nodes. ANNs with a single hidden layer node required up to 150 cycles for training.

Validation against direct binding

The prediction of binding peptides was also validated against the results of direct binding assays on a set of 16mer peptides from the tyrosine phosphatase IA-2 (Honeyman *et al.*, 1997). The experimental binding affinity was compared to prediction based on the highest scoring nonamer within each peptide (Table 5). All 916 nonamer peptides in the initial data set (which did not include any IA-2 peptides) were used to train the ANN.

In binary classification, all high-affinity binders, 82% of moderate-affinity binders, 30% of low-affinity binders and 70% of non-binders were correctly predicted. These results are similar to those from the internal cross-validation. There was a highly significant association between predicted and experimental binding (Kruskal–Wallis test: $P = 0.0001$).

Table 5. Experimental validation of the predictive performance of PERUN. Binding affinities are designated H, M, L and NB for high, moderate, low and zero affinity, respectively

Experimental	Predicted binding affinity			
	H	M	L	NB
H	0	1	1	0
M	3	4	2	1
L	0	2	1	7
NB	1	4	7	28

Comparison of PERUN with other prediction methods

PERUN was compared with the quantitative matrix and a weighted binding motif method, using the experimental binding affinities of IA-2 peptides. The relatively small number of test peptides (62) was insufficient to demonstrate a statistically significant difference in the performance of the three methods by ROC analysis. However, the results suggest that the predictive performance of PERUN is comparable to that of the quantitative matrix of Hammer *et al.* (1994a), and is likely to be better than that of the binding motif (Table 6).

Table 6. Comparison of the performance of the three prediction methods. The measure of performance is the area under the ROC curve (Aroc) with the standard error area given in parentheses. A value of Aroc = 0.5 indicates binary classification by random guessing, while Aroc = 1 indicates correct prediction for all test cases. Empirically, values of Aroc > 0.7 are considered as significant (Swets, 1988). The analysis was performed by comparing predictions at three arbitrarily defined thresholds for the definition of binding peptides. An ANN with two hidden layer nodes was used in prospective PERUN predictions

Prediction method	Aroc for arbitrary binding-definition threshold		
	Low affinity	Moderate affinity	High affinity
PERUN	0.73 (0.06)	0.86 (0.06)	0.88 (0.06)
MATRIX	0.73 (0.06)	0.82 (0.07)	0.87 (0.07)
MOTIF	0.63 (0.07)	0.69 (0.09)	0.74 (0.1)

Discussion

Our prime objective was to design a method for the prediction of MHC class II-binding peptides that could integrate experimental data and expert knowledge with the search and classification tools of the information science. The results indicate that we have largely succeeded in meeting this objective. PERUN predictions of peptide binding to HLA-DR4(B1*0401) are as good as or better than alternative methods. Furthermore, new peptides and their binding affinities can be incorporated simply into the primary data set, followed by the application of an EA to create a new training set. Thus, PERUN is adaptive, allowing theoretical predictions to be combined with experimentation in a two-way in-

terchange of information, refining both during the process. This represents a significant advantage in comparison with other methods which cannot be improved just by being used. ANNs can also be trained for specific requirements, e.g. high specificity or high sensitivity.

Allelic variants of HLA-DR molecules display high structural and functional similarity (Madden, 1995; Jardetzky *et al.*, 1996), and therefore this approach is likely to be generally applicable for predicting peptides that bind to other HLA-DR molecules.

An ANN as a classification device is well suited for extracting and learning peptide-MHC molecule binding rules, because it is adaptive, can generalize, deal with non-linear problems, and handle imperfect or incomplete data (Hammerstrom, 1993). More complex ANNs may perform better with increased number of available training data. A larger number of hidden layer nodes would cause the complexity of the ANN to exceed the available training data and probably result in a poorly defined learning problem (Amari *et al.*, 1995), with a tendency to memorize data rather than generalize and extract rules. The number of linearly separable regions in input space M , with J hidden layer nodes, is $M = 2^J$ (see Zurada, 1992, pp. 216–218). Therefore, the maximum number of classes for 1, 2, 3 and 4 hidden layer nodes is 2, 4, 8 and 16, respectively. The ANN with a single hidden node is therefore expected to perform well in two-class classification, i.e. discrimination between binders and non-binders; conversely, good performance in four-class classification requires two or more hidden layer nodes. Comparison of linear network to those with 1–4 hidden layer units showed no difference in performance. The possibilities which could explain this include: (i) peptide binding to the DR4 molecule is a linear problem that could be modeled by a single matrix or (ii) peptide binding to the DR4 molecule is non-linear, but available data are biased towards a linear model for historical reasons. Accumulation of binding data should help find an answer to the question of which model is appropriate. The recent findings indicate that binding of peptides to MHC class II molecules is a non-linear problem, influenced by both independent and inter-dependent binding of each amino acid within the peptide, and by other factors such as the overall structure of the peptide. This view is supported by crystallographic analysis (Jardetzky *et al.*, 1996). Raddrizzani *et al.* (1997) showed, experimentally, interdependence of individual amino acids on peptide binding to the HLA-DQ isotype of human class II MHC molecules. Therefore, the solution space for binders may comprise disjoint regions not encodable by a single matrix. The capacity of PERUN to cope with non-linear data is therefore essential for the prediction of peptide binding to the broad range of class II MHC molecules.

The quality of ANN prediction depends on the quality of training data as well as the complexity of the solution space. Of 338 binding peptides in the initial data set, 224 contained

two or more SAAR residues, resulting in a huge combinatorial space of possible alignments. Pre-processing was therefore a critical step because the task of selecting the most appropriate alignment is computationally complex. We used EA, the search method suitable for solving computationally difficult problems (Forrest, 1993), to align peptides. The rationale for accomplishing correct peptide alignment was to combine the power of EA with the realistic assumptions derived from available expert knowledge of anchor positions.

In an attempt to force as much divergence within the population of matrices as possible, we opted to exclude a cross-over operator which forces individuals to swap whole blocks of their genomes. The sensitivities of matrices of the final generation were ~60% and specificities were almost 100%. A matrix whose coefficients were calculated as the average of corresponding coefficients of the final generation matrices had a sensitivity of <30%, indicating that the individual matrices captured disjoint regions in the solution space. Preliminary EA experiments, including cross-over operator, resulted in a population of matrices of high similarity, promoting a linear model of peptide alignment. With the intention of using PERUN for the prediction of peptide binding to diverse class II MHC molecules, a non-linear model for peptide alignment has been preferred and the cross-over operator was excluded. The arbitrarily selected size of the population of matrices (10) appears to be sufficient for solving this alignment problem. The selection of the EA and ANN parameters provides a good balance between reasonable performance and the computational requirements.

There are several avenues for improvement of PERUN. They include: (i) addition of more experimental data; (ii) minimization of biases that are present in the data set by excluding some peptides; (iii) incorporation of additional conceptual knowledge, e.g. more refined anchor data; (iv) optimization of the network architecture; (v) introduction of alternative alignment methods; (vi) investigation of peptide representations that are potentially more appropriate for the ANN stage. An efficient means to improve prediction would be via an 'adaptive loop', to feed back results of experimental validation. Two kinds of experimental data can be used: new peptide sequences with their binding affinities or experimentally determined primary anchor positions within peptides. We did not use information on experimentally determined primary anchors within specific peptides, but this has the potential to capture more specific, refined rules for binding. The potential improvements in four-class classification must address one or more of the following: (i) a relatively small data set; (ii) an insufficient number of examples of low and moderate binding affinity; (iii) an arbitrary definition of class boundaries; (iv) the adequacy of one output node with sigmoid output for the classification of binding affinities.

Computer models will increasingly enable scientists to exploit experimental data optimally and plan experiments.

PERUN was implemented primarily to minimize the number of peptides required to be synthesized and tested as possible T-cell epitopes. Prediction of the peptides that bind to specific MHC molecules has implications for several areas of medicine, most obviously to vaccine development and the immunotherapy of autoimmune disease and cancer. Application of a prediction method may have different requirements. For example, high sensitivity, to identify all possible peptides, or high specificity to capture peptides that bind with the highest affinity. The inherent features of PERUN enable these requirements to be met simply by adjusting prediction thresholds.

Acknowledgements

The authors are grateful to Margaret Thompson for secretarial assistance, Mark W. Watson for technical support including installation and maintenance of the PlaNet package, and to Antony Eugoni for performing ROC analysis. This work was supported by the National Health and Medical Research Foundation of Australia and a Diabetes Interdisciplinary Research Program grant from the Juvenile Diabetes Foundation International.

References

- Adams, H.P. and Koziol, J.A. (1995) Prediction of binding to MHC class I molecules. *J. Immunol. Methods*, **185**, 181–190.
- Amari, S., Murata, N., Mueller, K.R., Finke, M. and Yang, H. (1995) Asymptotic statistical theory of overtraining and cross-validation. University of Tokyo Technical Report METR 95-06.
- Beale, R. and Jackson, T. (1990) *Neural Computing: An Introduction*. Adam Hilger, Bristol.
- Brusic, V., Rudy, G. and Harrison, L.C. (1994) Prediction of MHC binding peptides using artificial neural networks. In Stonier, R. and Yu, X.H. (eds), *Complex Systems: Mechanism of Adaptation*. IOS Press, Amsterdam, pp. 253–260.
- Brusic, V., Rudy, G., Kyne, A.P. and Harrison, L.C. (1996) MHCPEP—a database of MHC-binding peptides: update 1995. *Nucleic Acids Res.*, **24**, 242–244.
- Brusic, V., Schönbach, C., Takiguchi, M., Ciesielski, V. and Harrison, L.C. (1997) Application of genetic search in derivation of matrix models of peptide binding to MHC molecules. *ISMB*, 75–83.
- Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A., Lane, W.S. and Strominger, J.L. (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.*, **178**, 27–47.
- Cresswell, P. (1994) Assembly, transport, and function of MHC class II molecules. *Annu. Rev. Immunol.*, **12**, 259–293.
- Davenport, M.P., Shon, I.A.P.H. and Hill, A.V.S. (1995) An empirical method for the prediction of T-cell epitopes. *Immunogenetics*, **42**, 392–397.
- Engelhard, V.H. (1994) Structure of peptides associated with class I and class II MHC molecules. *Annu. Rev. Immunol.*, **12**, 181–207.
- Forrest, S. (1993) Genetic algorithms: principles of natural selection applied to computation. *Science*, **261**, 872–878.
- Gibbs, A.J. and McIntyre, G.A. (1970) The diagram, a method for comparing sequences. Its use with amino acids and nucleotide sequences. *Eur. J. Biochem.*, **16**, 1–11.
- Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading.
- Hammer, J., Valsasini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B. and Sinigaglia, F. (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. *Cell*, **74**, 197–203.
- Hammer, J., Bono, E., Gallazzi, F., Belunis, C., Nagy, Z. and Sinigaglia, F. (1994a) Precise prediction of MHC class II-peptide interaction based on peptide side chain scanning. *J. Exp. Med.*, **180**, 2353–2358.
- Hammer, J., Nagy, Z.A. and Sinigaglia, F. (1994b) Rules governing peptide-class II MHC molecule interactions. *Behring Inst. Mitt.*, **94**, 124–132.
- Hammerstrom, D. (1993) Neural networks at work. *IEEE Spectrum*, **30**, 26–32.
- Harrison, L.C., Honeyman, M.C., Trembleau, S., Gregori, S., Gallazzi, F., Augstein, P., Brusic, V., Hammer, J. and Adorini, L. (1997) A peptide-binding motif for I-A^{B7}, the class major histocompatibility (MHC) molecule of NOD and Biozzi AB/H mice. *J. Exp. Med.*, **185**, 1013–1021.
- Holland, J.H. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.
- Honeyman, M.C., Brusic, V. and Harrison, L.C. (1997) Strategies for identifying and predicting islet autoantigen T-cell epitopes in insulin-dependent diabetes (IDDM). *Ann. Med.*, **29**, 401–404.
- Jardetzky, T.S., Brown, J.H., Gorga, J.C., Stern, L.C., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1996) Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc. Natl Acad. Sci. USA*, **93**, 734–738.
- Kondo, A. et al. (1995) Prominent roles of secondary anchor residues in peptide binding to HLA-A24 human class molecules. *J. Immunol.*, **155**, 4307–4312.
- Madden, D.R. (1995) The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol.*, **13**, 587–622.
- McClelland, T.L. and Rumelhart, D.E. (1986) *Parallel Distributed Processing*. MIT Press, Cambridge.
- Miyata, J. (1991) *A User's Guide to PlaNet Version 5.6*. Computer Science Department, University of Colorado, Boulder, CO.
- Nijman, H.W., Houbiers, J.G., Vierboom, M.P., van der Burg, S.H., Drijfhout, J.W., D'Amato, J., Kenemans, P., Melief, C.J. and Kast, W.M. (1993) Identification of peptide sequences that potentially trigger HLA-A2.1-restricted cytotoxic lymphocytes. *Eur. J. Immunol.*, **23**, 1547–1553.
- Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163–175.
- Radrizzani, L., Sturniolo, T., Guenot, J., Bono, E., Galazzi, F., Nagy, Z.A., Sinigaglia, F. and Hammer, J. (1997) Different modes of peptide interaction enable HLA-DQ and HLA-DR molecules to bind diverse peptide repertoires. *J. Immunol.*, **159**, 703–711.
- Rammensee, H.G., Falk, K. and Rotzschke, O. (1993) Peptides naturally presented by MHC class I molecules. *Annu. Rev. Immunol.*, **11**, 213–244.

- Rammensee, H.G., Friede, T. and Stevanovic, S. (1995) MHC ligands and peptide motifs: first listing. *Immunogenetics*, **41**, 178–228.
- Rothbard, J.B., Marshall, K., Wilson, K.J., Fugger, L. and Zaller, D. (1994) Prediction of peptide affinity to HLA DRB1*0401. *Int. Arch. Allergy Immunol.*, **105**, 1–7.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning representations by back-propagating errors. *Nature*, **323**, 533–536.
- Schönbach, C., Ibe, M., Shiga, H., Takamiya, Y., Miwa, K., Nokiara, K. and Takiguchi, M. (1995) Fine tuning of peptide binding to HLA-B*3501 molecules by nonanchor residues. *J. Immunol.*, **154**, 5951–5958.
- Sette, A., Buus, S., Appella, E., Smith, J.A., Chesnut, R., Miles, C., Colon, S.M. and Grey, H.M. (1989) Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc. Natl Acad. Sci. USA*, **86**, 3296–3300.
- Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Weiss, S.M. and Kulikowski, C.A. (1990) *Computer Systems that Learn*. Morgan Kaufman Publishers Inc., San Mateo.
- Zurada, J.M. (1992) *Introduction to Artificial Neural Systems*. West Publishing Company, St Paul.



Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for

Limits Preview/Index History Clipboard Details

Abstract Text

About Entrez

Text Version

1: Hum Immunol. 1997 Nov;58(1):1-11.

Related Article

Entrez PubMed

Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services

Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Related Resources

Order Documents
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

ELSEVIER SCIENCE
FULL-TEXT ARTICLE

A structure-based algorithm to predict potential binding peptide-MHC molecules with hydrophobic binding pockets.

Altuvia Y, Sette A, Sidney J, Southwood S, Margalit H.

Department of Molecular Genetics and Biotechnology, Hebrew University-Hadassah Medical School, Jerusalem, Israel.

Binding of peptides to MHC class I molecules is a prerequisite for their recognition by cytotoxic T cells. Consequently, identification of peptides that will bind to a given MHC molecule must constitute a central part of any algorithm for prediction of T-cell antigenic peptides based on the amino acid sequence of the protein. Binding motifs, defined by anchor positions only, have proven to be insufficient to ensure binding, suggesting that other positions along the peptide sequence also affect peptide-MHC interaction. The second phase of prediction schemes therefore takes into account the effect of all positions along the peptide sequence, and is based on position-dependent-coefficients that are used in the calculation of a peptide score. These coefficients can be extracted from a large ensemble of binding sequences that were tested experimentally, or derived from structural considerations, as in the algorithm developed by us recently. This algorithm uses the coordinates of solved complexes to evaluate the interactions of peptide acids with MHC contact residues, and results in a peptide score that reflects its binding energy. Here we present our analysis for peptide binding to four MHC alleles (HLA-A2, HLA-A68, HLA-B27 and H-2Kb), and compare the prediction of the algorithm to experimental binding data. The algorithm performs successfully in predicting peptide binding to MHC molecules with hydrophobic binding pockets but not when MHC molecules with hydrophilic, charged pockets are considered. For MHC molecules with hydrophobic pockets it is demonstrated how the algorithm succeeds in distinguishing binding from non-binding peptides, and in the high ranking of immunogenic peptides within all overlapping same-length peptides spanning their respective protein sequences. The latter property of the algorithm makes it a useful tool in the rational design of peptide vaccines aimed at T-cell immunity.

PMID: 9438204 [PubMed - indexed for MEDLINE]

EXHIBIT

D